

# Targeted Amplicon Sequencing for Single-Nucleotide-Polymorphism Genotyping of Attaching and Effacing *Escherichia coli* O26:H11 Cattle Strains via a High-Throughput Library Preparation Technique

Sarah A. Ison,<sup>a</sup> Sabine Delannoy,<sup>b</sup> Marie Bugarel,<sup>a</sup> Tiruvoor G. Nagaraja,<sup>c</sup> David G. Renter,<sup>c</sup> Henk C. den Bakker,<sup>a</sup> Kendra K. Nightingale,<sup>a</sup> Patrick Fach,<sup>b</sup> Guy H. Loneragan<sup>a</sup>

Department of Animal and Food Sciences, Texas Tech University, Lubbock, Texas, USA<sup>a</sup>; ANSES Food Safety Laboratory, Platform IdentityPath, Université Paris-Est, Maisons-Alfort, France<sup>b</sup>; Department of Diagnostic Medicine Pathobiology, Kansas State University, Manhattan, Kansas, USA<sup>c</sup>

**Enterohemorrhagic *Escherichia coli* (EHEC) O26:H11, a serotype within Shiga toxin-producing *E. coli* (STEC) that causes severe human disease, has been considered to have evolved from attaching and effacing *E. coli* (AEEC) O26:H11 through the acquisition of a Shiga toxin-encoding gene. Targeted amplicon sequencing using next-generation sequencing technology of 48 phylogenetically informative single-nucleotide polymorphisms (SNPs) and three SNPs differentiating Shiga toxin-positive (*stx*-positive) strains from Shiga toxin-negative (*stx*-negative) strains were used to infer the phylogenetic relationships of 178 *E. coli* O26:H11 strains (6 *stx*-positive strains and 172 *stx*-negative AEEC strains) from cattle feces to 7 publically available genomes of human clinical strains. The AEEC cattle strains displayed synonymous SNP genotypes with *stx*<sub>2</sub>-positive sequence type 29 (ST29) human O26:H11 strains, while *stx*<sub>1</sub>, ST21 human and cattle strains clustered separately, demonstrating the close phylogenetic relatedness of these Shiga toxin-negative AEEC cattle strains and human clinical strains. With the exception of seven *stx*-negative strains, five of which contained *espK*, three *stx*-related SNPs differentiated the STEC strains from non-STEC strains, supporting the hypothesis that these AEEC cattle strains could serve as a potential reservoir for new or existing pathogenic human strains. Our results support the idea that targeted amplicon sequencing for SNP genotyping expedites strain identification and genetic characterization of *E. coli* O26:H11, which is important for food safety and public health.**

Shiga toxin-producing *Escherichia coli* (STEC) O26 strains have been associated with large food-borne outbreaks worldwide and can result in severe human illness (1–3). Zoonotic pathogens, such as STEC O26, can contaminate produce, meat, and dairy products through the transmission from ruminants, such as cattle (4, 5). Enterohemorrhagic *E. coli* (EHEC) O26:H11, a subgroup within STEC that causes severe human disease, is believed to have originated from attaching and effacing *E. coli* (AEEC) O26:H11 through the acquisition of a Shiga toxin-encoding gene (6). AEEC strains cause typical attaching and effacing lesions in the intestinal mucosa and are characterized by the absence of a Shiga toxin gene (*stx*) and the presence of an intimin gene (*eae*), carried on the locus of enterocyte effacement (LEE) (6, 7).

A new highly pathogenic EHEC O26:H11 clone containing only *stx*<sub>2a</sub> has previously been described, with an increased virulence and disease severity compared to O26:H11 strains containing *stx*<sub>1</sub> alone or in combination with *stx*<sub>2</sub> (8). Since its emergence in Germany in the early 1990s, this new *stx*<sub>2a</sub> clone has disseminated around the world (1, 4, 8–18). Multilocus sequence typing has shown that this clone belongs to sequence type 29 (ST29) (8), which is associated with severe human disease, while the phylogenetically related group ST21 is associated with less-severe disease (6, 10, 19).

More recently, a new subpopulation of pathogenic O26:H11 strains harboring *stx*<sub>2a</sub> or *stx*<sub>2d</sub> belonging to ST29, has been identified in hemolytic-uremic syndrome (HUS) patients in France (19). These *stx*<sub>2</sub>-positive strains differ from the characteristic typical O26:H11 EHEC strains (20). Indeed, while they possess a *stx* gene and the *eae*-beta allele, they lack other genetic features, such as the effector translocated by the type III secretion system, *espK*,

and contain a different allele of the aerobic respiratory control protein A (*arcA*) (19). They also exhibit a distinct clustered regularly interspaced short palindromic repeat (CRISPR) locus. To detect this French O26:H11 subpopulation, the CRISPR PCR marker SP\_O26-E was developed (19). Surprisingly, AEEC O26:H11 strains of bovine fecal origin isolated in the United States have been observed to contain virulence-associated genes and the CRISPR marker SP\_O26-E that is specific for the detection of the new French clone (21).

The evolutionary phylogenetic relationship of EHEC O26:H11/H<sup>-</sup> has been investigated on strains of European origin employing whole-genome sequencing (WGS) in which four distinct clonal complexes (CCs) were observed using 48 phylogenetically informative single-nucleotide polymorphisms (SNPs) (22). The highly virulent German O26:H11 *stx*<sub>2a</sub> clone was identified in a

Received 4 October 2015 Accepted 4 November 2015

Accepted manuscript posted online 13 November 2015

Citation Ison SA, Delannoy S, Bugarel M, Nagaraja TG, Renter DG, den Bakker HC, Nightingale KK, Fach P, Loneragan GH. 2016. Targeted amplicon sequencing for single-nucleotide-polymorphism genotyping of attaching and effacing *Escherichia coli* O26:H11 cattle strains via a high-throughput library preparation technique. *Appl Environ Microbiol* 82:640–649. doi:10.1128/AEM.03182-15.

Editor: C. A. Elkins

Address correspondence to Sarah A. Ison, sarah.hazenfield@ttu.edu.

Supplemental material for this article may be found at <http://dx.doi.org/10.1128/AEM.03182-15>.

Copyright © 2016 Ison et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 Unported license](https://creativecommons.org/licenses/by/4.0/).

single CC, different from the hitherto described strains (22). To distinguish between strains containing a Shiga toxin-encoding gene and those strains lacking a Shiga toxin gene, the O-antigen gene cluster in *E. coli* O26:H11 has been investigated (23). Three SNPs were identified that, when used in combination, detected all *stx*-positive strains; one SNP captured a subset of strains containing only *stx*<sub>2</sub> (23).

High-throughput next-generation sequencing (NGS) technologies have created a gateway into genomic research through accelerated genetic discoveries. WGS has been shown to be a powerful and informative tool for SNP discovery in which markers that provide genetic variations for hypothesis investigation are identified (22, 24–27). Although performing WGS on individual bacterial strains is becoming more cost-effective (28–30), sequencing many strains in the discovery process may be unnecessary (24), and computational genomics and bioinformatics pipelines can create a bottleneck in data analysis (24, 31). To circumvent these issues, targeted amplicon sequencing of SNPs using NGS can be used as a powerful approach for the rapid phylogenetic classification and characterization of several hundred to thousands of bacterial strains.

Extensive investigation into the phylogeny of STEC O26:H11 strains has been performed (22, 26, 32); however, there are missing links in how Shiga toxin-negative, *eae*-positive AEEC strains are related to pathogenic O26:H11 strains. Here we describe the investigation of the phylogenetic relationship and genetic diversity of AEEC O26:H11 bovine strains by the use of a high-throughput method for preparing targeted SNP amplicons for NGS.

## MATERIALS AND METHODS

**Bacterial strains.** *E. coli* O26:H11 bovine fecal strains ( $n = 178$ ), isolated from May to July 2011 in cattle originating from a commercial feedlot in the United States housing 17,148 cattle in 40 pens (33), were investigated. These strains have previously been studied for their association with select virulence-associated and O26 serogroup-specific molecular markers along with the amplification of their CRISPR 1 and 2a loci. These strains displayed 37 unique genetic profiles, diversity types (DTs), determined by the combination of molecular markers and CRISPR alleles (Table 1) (21). Out of the strains investigated, 168 consisted of Shiga toxin-negative and *eae*-beta-positive strains. This combination of characteristics classifies these strains as AEEC (6, 7). Of these strains, 161 displayed the characteristic profile of the recently identified new French *stx*<sub>2a</sub> and *stx*<sub>2d</sub> clone containing the CRISPR PCR marker SP\_O26-E (DTs 9 to 37) (19, 21). An additional five strains were Shiga toxin positive, *eae*-beta positive, and *espK* positive, classifying them as EHEC O26:H11 (DTs 2 and 3). Also, five strains were Shiga toxin negative, *eae*-beta positive, and *espK* positive (DT 4), classifying them as EHEC-like O26:H11. One strain was *stx*<sub>1</sub> positive and *espK* negative and contained SP\_O26-E (DT 1) (21).

**Targeted single-nucleotide polymorphisms.** Oligonucleotides were designed to target 51 previously identified SNPs (22, 23). Forty-eight SNPs, previously determined to be phylogenetically informative by Bletz et al. (22), were investigated in the bovine O26:H11 strains to determine the phylogenetic association of these strains to EHEC O26 strains of human origin that group into four distinct SNP clonal complexes (22). Three additional SNPs, located in the O-antigen gene cluster, which had been determined to differentiate Shiga toxin-positive strains from Shiga toxin-negative strains were investigated (23). A combination of STEC allele *rmlA* or *fnl-1* with *wzx* was observed to capture all *stx*-positive O26 strains with *wzx* capturing a subset of strains that contained only *stx*<sub>2</sub> (23).

SNPs with reference SNP genotypes are provided in Table S1 in the supplemental material. *E. coli* O26:H11 strain 11368 NCBI reference sequence NC\_013361.1 (GenBank accession no. AP010953.1) was used as the reference strain (34).

**Target-specific primer characteristics.** The Fluidigm Access Array system (Fluidigm, South San Francisco, CA) provides an expedited workflow from strain to targeted amplicons that are sequence ready and compatible on multiple NGS platforms. The Access Array two-primer target-specific PCR amplification generates 2,304 PCRs on a single integrated fluidic circuit (IFC) through the combination of 48 samples with 48 target-specific primer pairs (35). The *Access Array System for Illumina Sequencing Systems User Guide* (36) was used with the modifications described below to perform the two-primer target-specific PCR amplification on the 48.48 Access Array integrated fluidic circuit. Targeted next-generation sequencing with the Fluidigm Access Array amplifies a short genomic region of 100 to 150 bp in length with a final product of 150 to 200 bp. Oligonucleotides were designed to target the absolute genome positions of the previously identified 51 SNPs with the targeted SNP to be placed in the center of the final amplified product (see Table S2 in the supplemental material). Oligonucleotides were designed with a melting temperature of 59 to 62°C and optimum temperature of 60°C. Primers were selected with required specifications using Primer3Plus (37) and OligoCalc, an oligonucleotide properties calculator (38).

Common sequence 1 (CS1) (5'-ACACTGACGACATGGTTCTACA-3') and common sequence 2 (CS2) (5'-TACGGTAGCAGAGACTTGGTCT-3') universal primer sequences, required for Illumina MiSeq amplicon tagging and indexing, were added to the 5' ends of all target-specific forward (TSF) and reverse (TSR) primers, respectively. The *Fluidigm Access Array - Generate Tagged Primers Workbook* (39) was used to add universal sequence tags to the target-specific primers (see Table S3 in the supplemental material). Primers were purchased from Eurofins Genomics (Ebersberg, Germany).

**Primer validation.** Target-specific primer validation for two-primer reactions on the 48.48 Access Array IFC was modified to include target-specific primers and was performed in a 96-well PeqSTAR thermocycler. Two *E. coli* O26:H11 strains and water were used as positive and negative controls for primer validation, respectively. The validation master mix solution with final concentrations contained 1× FastStart high-fidelity reaction buffer without MgCl<sub>2</sub> (Roche, Meylan, France), 4.5 mM MgCl<sub>2</sub> (Roche), 5% dimethyl sulfoxide (DMSO) (Roche), 200 μM each PCR-grade nucleotide mix (Roche), 1× Access Array loading reagent (Fluidigm, South San Francisco, CA), tagged TSF and TSR primers (200 nM each), 10 ng/μl DNA, 0.05 U/μl FastStart high-fidelity enzyme blend (Roche), molecular grade water to a final volume of 10 μl. The thermocycler program run with 35 cycles was performed as recommended in the Fluidigm targeted sequencing primer validation protocol (36) and provided in Table S4 in the supplemental material. PCR products were checked for each PCR using the FlashGel system (Lonza, France) with the FlashGel DNA 50-bp to 1.5-kb high-concentration marker to ensure that only one PCR product of the correct size was generated for each target.

Three duplex primer pairs were validated as previously described with substitution of 1 μl of molecular grade water with 0.5 μl of each additional primer. Primer pairs selected for duplex were first compared on the Thermo Fisher Scientific multiple primer analyzer (<https://www.thermofisher.com/us/en/home/brands/thermo-scientific/molecular-biology/molecular-biology-learning-center/molecular-biology-resource-library/thermo-scientific-web-tools/multiple-primer-analyzer.html>). Duplexing primer pairs allowed 51 targeted SNPs to be run on a single 48.48 Access Array IFC for a total of 2,448 amplicons generated on each 48.48 Access Array IFC. Primers combined in duplex for targeting SNPs were *fruR* with *murC*, *ybcF* with *sdhB*, and *ybjG* with *hyaE*.

**48.48 Access Array IFC workflow.** The two-primer target-specific PCR amplification was performed at the iGenSeq Genotyping and Sequencing Core facility at the ICM Brain and Spine Institute (Paris, France) on the 48.48 Access Array IFC according to the manufacturer's recommendations (40). The Fluidigm 48.48 Access Array workflow for Illumina sequencing involves two independent PCRs as follows (Fig. 1). First, the 48.48 Access Array IFC is primed on a pre-PCR IFC controller AX (Fluidigm), 48 separate primer solutions are added to individual primer inlets,

TABLE 1 Single-nucleotide polymorphism clonal complexes, Shiga toxin-associated alleles, and molecular characteristics of cattle O26:H11 strains

DT <sup>a,b</sup>	No. of isolates	SNP CC <sup>c</sup>	STEC allele <sup>d</sup>	SNP in the following gene <sup>e</sup> :			Presence or absence of the following gene:					CRISPR1 allele <sup>a</sup>	CRISPR2a allele <sup>a,e</sup>	CT <sup>a,f</sup>
				<i>rmlA</i> STEC (G → T)	<i>wzx</i> STEC (T → G)	<i>fli-1</i> STEC (G → A)	<i>stx</i> <sub>1</sub> <sup>a</sup>	<i>espK</i> <sup>a</sup>	<i>ehxA</i> <sup>a</sup>	<i>arcA2</i> <sup>a</sup>	SP_O26-E <sup>a</sup>			
1	1	1	0	G	T	G	+	-	-	-	+	29	67	3
2	3	3	1	<b>T</b>	T	<b>A</b>	+	+	+	+	-	11	160	1
3	2	4	1	<b>T</b>	T	<b>A</b>	+	+	+	+	-	9	159	2
4	5	1	1	G	<b>G</b>	G	-	+	+	+	-	29	4	4
5	1	1	0	G	T	G	-	-	-	-	-	U	24	18
6	1	1	0	G	T	G	-	-	-	-	-	29	162	6
7	3	1	0	G	T	G	-	-	-	-	-	131	72	14
8	1	1	0	G	T	G	-	-	-	-	-	29	67	3
9	7	1	0	G	T	G	-	-	-	-	+	29	67	3
10	1	1	0	G	T	G	-	-	-	-	+	116	U-4	22
11	8	1	0	G	T	G	-	-	-	-	+	116	70	13
12	2	1	1	G	<b>G</b>	G	-	-	-	-	+	29	67	3
12	62	1	0	G	T	G	-	-	-	-	+	29	67	3
13	3	1	0	G	T	G	-	-	-	-	+	130	67	17
14	1	1	0	G	T	G	-	-	-	-	+	29	U-5	23
15	1	1	0	G	T	G	-	-	-	-	+	29	U-6	24
16	2	1	0	G	T	G	-	-	-	-	+	29	67	3
17	2	1	0	G	T	G	-	-	-	-	+	29	67	3
18	1	1	0	G	T	G	-	-	-	-	+	29	67	3
19	14	1	0	G	T	G	-	-	-	-	+	11	70	5
20	13	1	0	G	T	G	-	-	-	-	+	125	161	7
21	8	1	0	G	T	G	-	-	-	-	+	126	161	8
22	1	1	0	G	T	G	-	-	-	-	+	116	70	13
23	1	1	0	G	T	G	-	-	-	-	+	129	72	15
24	1	1	0	G	T	G	-	-	-	-	+	129	70	16
25	1	1	0	G	T	G	-	-	-	-	+	11	70	5
26	9	1	0	G	T	G	-	-	-	-	+	125	161	7
27	2	1	0	G	T	G	-	-	-	-	+	126	161	8
28	1	1	0	G	T	G	-	-	+	-	+	126	161	8
29	1	1	0	G	T	G	-	-	-	-	+	126	161	8
30	3	1	0	G	T	G	-	-	-	-	+	120	67	9
31	1	1	0	G	T	G	-	-	-	-	+	120	159	10
32	3	1	0	G	T	G	-	-	-	-	+	120	24	11
33	1	1	0	G	T	G	-	-	-	-	+	124	160	12
34	7	1	0	G	T	G	-	-	-	-	+	120	U-1	19
35	1	1	0	G	T	G	-	-	-	-	+	120	U-2	20
36	1	1	0	G	T	G	-	-	-	-	+	124	U-3	21
37	1	1	0	G	T	G	-	-	+	-	+	120	U-2	20

<sup>a</sup> These *E. coli* O26:H11 strain characteristics originated from previously published results (21) which investigated the diversity and molecular characteristics of the strains in this study.

<sup>b</sup> The diversity type (DT) classification previously reported for the O26:H11 strains was determined through the combination of molecular markers and CRISPR alleles (21).

<sup>c</sup> SNP clonal complex (SNP CC) for which *E. coli* O26:H11 strains were classified based on the 48 phylogenetically informative SNPs previously identified in the SNP discovery study (22).

<sup>d</sup> Allele pattern of O-antigen genes for the presence of either the *stx*<sub>1</sub> or *stx*<sub>2</sub> STEC-associated allele.

<sup>e</sup> CRISPR alleles denoted with a U indicate that the CRISPR amplicon was unable to be amplified.

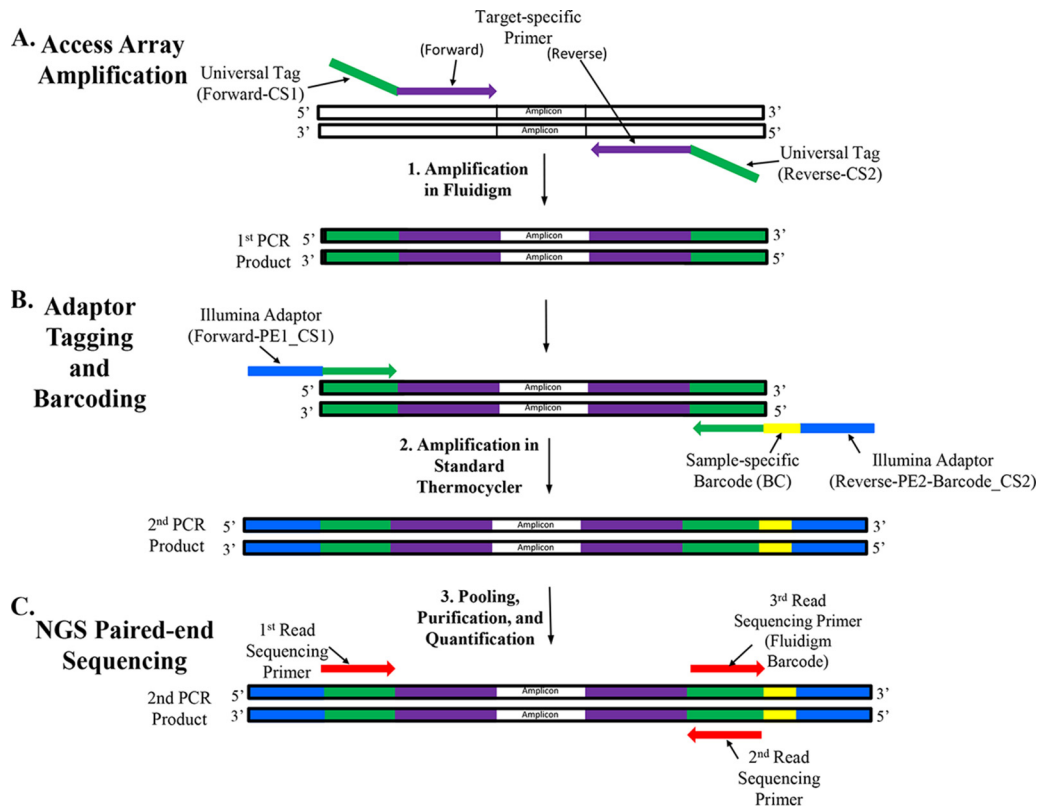
<sup>f</sup> Each unique combination of CRISPR1 and CRISPR2a alleles was assigned a CRISPR type (CT).

<sup>g</sup> Right arrows indicate the nucleotide SNP change; letters in bold indicate that the STEC allele is displayed.

and 48 DNA sample mix solutions are added to individual sample inlets. The pre-PCR IFC controller AX combines each of the 48 primers and 48 samples into 2,304 separate microreaction chambers. The 48.48 Access Array IFC is then placed onto the FC1 cycler (Fluidigm) to complete the first thermocycler reaction (Fig. 1A). After thermal cycling, an Access Array harvest solution is added to the Access Array IFC and moved to a post-PCR IFC controller AX to perform the Harvest (151×) run script. In this step, all the amplicons generated for each sample are collected in the corresponding sample inlet. The tagged target-specific PCR amplicons are

then transferred from the Access Array IFC to a 96-well PCR plate, resulting in sample-specific pooled amplicons. Four 48.48 Access Array IFCs were used in this study. For proper loading of the Access Array IFC, water was added in place of DNA in sample solution to combine with primer solution for incomplete sample plates.

In the second PCR, performed on a conventional thermocycler, Illumina sequence-specific adaptors with individual sample barcodes on the reverse primer were attached to the amplicons produced in the first PCR (39) (Fig. 1B). The Fluidigm Access Array barcode library for Illumina



**FIG 1** 48.48 Access Array high-throughput workflow. (A) The first amplification on the Fluidigm system resulting in targeted amplicons containing universal tags on both ends. (B) A standard thermocycler is used to perform the second PCR, adding Illumina-specific adaptors and a sample-specific barcode. (C) Paired-end next-generation sequencing (NGS) is performed incorporating first-read, second-read, and third-read sequencing primers to read the sample-specific barcode.

sequencers 384 (single-direction) kit (PN 100-4876; Fluidigm) barcode numbers 1 to 178 were used for amplicon tagging. The Illumina-specific primer sequences are as follows: forward (PE1\_CS1), 5'-AATGATACGG CGACCACCGAGATCTACTGACGACATGGTTCTACA-3'; and reverse (PE2\_Barcode\_CS2), 5'-CAAGCAGAAGACGGCATAACGAGAT-(barcode)-TACGGTAGCAGAGACTTGGTCT-3'. Each DNA sample was assigned a unique barcode sequence, which is used for demultiplexing samples downstream.

The PCR products generated were qualified using a LabChip Gx reader (PerkinElmer Inc.) (see Fig. S1A and S1B in the supplemental material). All PCR products were then pooled to create a single product library. The quality of the pooled library was checked on a LabChip GX reader (PerkinElmer Inc.), and the library concentration was determined on the QuantiFluor system (Promega) (Fig. S1C).

**Amplicon sequencing using the Illumina MiSeq.** Custom Access Array primers (FL1 and FL2) are incorporated to read the amplicons generated from the Access Array IFC. FL1 and FL2 primer mixes were prepared according to the *Access Array System for Illumina Sequencing Systems User Guide* (36). The FL1 primer mix was prepared by combining CS1 oligonucleotide (5'-A+CA+CTG+ACGACATGGTTCTACA-3') and CS2 oligonucleotide (5'-T+AC+GGT+AGCAGAGACTTGGTCT-3'). The FL2 primer mix combines the CS1rc oligonucleotide (5'-T+GT+AG+A ACCATGTCGTCAGTGT-3') and CS2rc oligonucleotide (5'-A+GAC+CA+AGTCTCTGCTACCGTA-3'); locked nucleic acid (LNA) nucleotides are preceded by a + sign; oligonucleotides were synthesized by TIB MOLBIOL (Berlin, Germany).

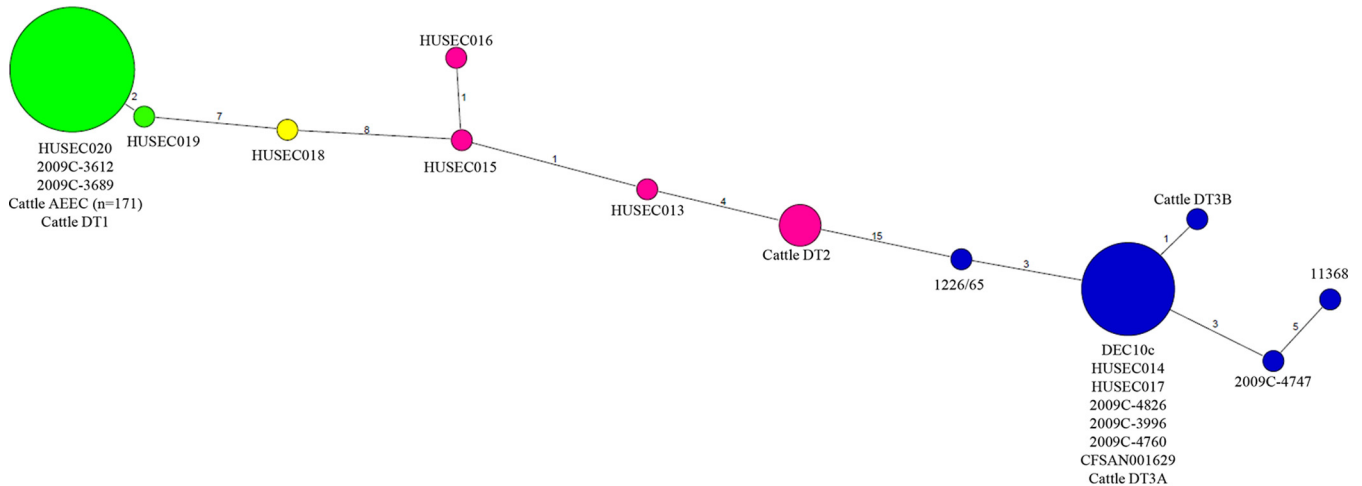
Denaturation and dilution of the Access Array pooled library was performed according to the Illumina preparing libraries for sequencing on the MiSeq protocol (41). The library was first diluted to 4 nM based on the

QuantiFluor concentration and the average length (in base pairs) of the libraries (calculation provided in Table S5 in the supplemental material). The denatured DNA for 4 nM library procedure was performed and diluted to a final concentration of 8 pM.

The denatured and diluted sample library was sequenced on an Illumina MiSeq instrument as a two 150-bp paired-end runs (MiSeq flow cell, v2 reagents) according to the *Illumina MiSeq System User Guide* (42) (Fig. 1C) with the following modifications. Sequencing was performed with an index read length of 10 cycles to read all 10 bases of the Fluidigm indexes (barcodes). If only 96 of the Fluidigm barcodes are used, i.e., 96 samples, only the first 8 bases of the indexes would need to be read.

**Bioinformatics analysis.** Sequencing reads were demultiplexed according to the sample-specific barcodes on the Illumina MiSeq sequencer. A reference sequence was created by concatenating the sequences of the 51 amplicons from the reference *stx*<sub>1</sub>-positive strain 11368 NCBI reference sequence NC\_013361.1 (GenBank accession no. AP010953.1), and sequence reads for each *E. coli* strain were mapped to this reference sequence using the CLC Genomics Workbench software version 7.5.1 (Qiagen, Aarhus, Denmark). Sequences were trimmed to ensure removal of Access Array barcodes with a quality score of 0.05 and a minimum read length of 15. The Basic Variant Detection tool in the CLC Genomics Workbench was used to identify and output SNP variants present. An R script (version 3.0.2) was written to reformat the CLC output of SNP variants and concatenate nucleotides at each SNP position for analysis. R script has been made publically available at <https://github.com/sarahannison>.

A minimum spanning tree was created by concatenating the 48 SNPs proposed by Bletz et al. (22) to identify the relationship of our Shiga toxin-negative cattle strains to four previously observed, clinically and phylogenetically associated, clonal complexes (CCs). Ten unique SNP



**FIG 2** Minimum spanning tree of 48 SNPs used to identify four clonal complexes and relationship of cattle Shiga toxin-negative O26:H11 strains. Clonal complex SNPs were previously proposed by Bletz et al. (22). Five Shiga toxin-positive cattle O26:H11 strains are included for reference (21). The four SNP clonal clusters (SNP CCs) are represented by different colors as follows: SNP CC1 is green, SNP CC2 is yellow, SNP CC3 is pink, and SNP CC4 is blue. The small numbers on the lines connecting the circles represent the differing number of SNPs between two nodes. The number of isolates present within a node are reflected by the size of the circle, with the exception that the graphical representation of the AEEC cattle O26:H11 node is presented at 5% of the actual size. In the Bletz minimum spanning tree (shown in Fig. 2 of reference 22), HUSEC015 and HUSEC016 are displayed in the same node; however, they differ in one SNP, creating a new node. This figure was created with Splitree (version 4.13.1) (43) characters > Jukes Cantor distances > MinSpanning Network for Distances.

profiles from 120 EHEC O26 isolates, which led to the designation of four CCs, were observed in the Bletz et al. SNP discovery study (22), and a representative strain was designated for each observed SNP genotype. The SNP genotypes of these 10 strains, 1226/65, 1557/77 (DEC10c), 2245/98 (HUSEC013), 5080/97 (HUSEC014), 126814/98 (HUSEC015), 5028/97 (HUSEC016), 3319/99 (HUSEC017), 1530/99 (HUSEC018), 1588/98 (HUSEC019), and 3271/00 (HUSEC020) (22), were used to guide the construction of the tree and determine the relationship of the cattle O26:H11 strains in this study to the four previously identified CCs. The minimum spanning tree was created in Splitree (43) (version 4.13.1) with character inputs, Jukes Cantor, and MinSpanning Network for Distances.

**SNP analysis of O26:H11 human clinical strains.** Publicly available whole-genome sequences of seven human clinical O26:H11 strains, originating from the United States, were investigated for the 48 phylogenetically informative SNPs and the three O-antigen STEC allele SNPs. Strains included 2009C-3612 (NCBI accession no. JHGZ000000000), 2009C-3689 (GenBank accession no. JHGX000000000), 2009C-3996 (GenBank accession no. JHGV000000000), 2009C-4747 (GenBank accession no. JHGM000000000), 2009C-4760 (GenBank accession no. JH GK000000000), 2009C-4826 (GenBank accession no. JHGI000000000), (44), and CFSAN001629 (GenBank accession no. AMXO000000000) (45). The Shiga toxin gene present (<https://www.patricbrc.org/>) (46) and multilocus sequence type (MLST) of each strain was determined (<http://mlst.warwick.ac.uk/mlst/dbs/Ecoli>) (47). The reference nucleotide sequence containing each SNP was aligned with the publicly available whole-genome sequence for each strain by using a local BLAST (version 2.2.29) (48) script (BLOPv2) that was developed in-house. This script utilizes a single text file containing all of the targeted nucleotide sequences in individual Fasta format and a local BLAST script to identify the degree of similarity with each sequence provided into a single .html output file. As a result, the nucleotide at each SNP position was quickly identified.

## RESULTS

**SNP genotyping by next-generation sequencing.** Of the 178 *E. coli* O26:H11 bovine strains that were investigated for the 48 phylogenetically informative SNPs and 3 Shiga toxin-associated SNPs, one strain, strain 4850A (see Fig. S1B in the supplemental material), did not generate the targeted library size and was ex-

cluded from the analysis. Qualification of the strain library is shown in Fig. S1C. The Illumina MiSeq platform sequencer produced 150-bp paired-end reads yielding  $3.0 \times 10^6$  total reads with approximately 16,949 reads per sample and 330 reads per targeted locus. Of the 9,027-targeted loci, 61 loci failed to be amplified or only one read for specific loci was produced (Table S6). Of the failed loci, 22 and 21 were observed in the *murC* and *rcsA* loci, respectively. Only two O26:H11 samples had two nondetermined loci, and no more than one locus was undetermined in all other samples.

**Classification of AEEC O26:H11 bovine strains according to phylogenetically informative SNPs.** Concatenated SNP genotypes of the 48 phylogenetically informative SNPs were created for all cattle O26:H11 strains, and a minimum spanning tree that included the reference strains (22) for each clonal complex was then constructed (Fig. 2). SNP clonal complex 1 (CC-1) contained strains that were either *stx*<sub>1</sub> ST29 or *stx*<sub>2</sub> positive and belonged to ST29 or a single-locus variant of ST29, ST396. It has previously been identified that a large majority of these AEEC O26:H11 cattle strains also belong to ST29 (21). Notably, a few *stx*<sub>2a</sub> ST29 strains that resulted in HUS also clustered in CC-1 in the original SNP discovery study (22). Analysis of the 48 phylogenetic informative SNPs determined that all of the AEEC O26:H11 strains ( $n = 171$ ) displayed the *stx*<sub>2a</sub> ST396 HUSEC020 strain SNP genotype (see Table S7 in the supplemental material) and clustered into CC-1, shown in green in Fig. 2. Thus, the bovine AEEC O26:H11 strains studied here appear related to this cluster of strains. The *stx*-positive, SP\_O26-E-positive strain, displaying the molecular marker and CRISPR allele combination for diversity type 1 (DT 1) (21) (Table 1) also displayed this CC-1 SNP genotype.

Three *stx*-positive strains, belonging to DT 2, clustered within CC-3 (shown in pink in Fig. 2), exhibited a four-SNP difference from strain HUSEC013, and differed in 15 SNPs from CC-4 strain 1226/65. One DT 3, *stx*-positive strain displayed the CC-4 (shown in blue in Fig. 2) SNP genotype synonymous with strains DEC10c,

HUSEC014, and HUSEC017. The other DT 3 strain, also grouped in CC-4, differed from this SNP profile at one SNP position.

**Nucleotide polymorphisms discriminating Shiga toxin-positive O26:H11 strains from Shiga toxin-negative O26:H11 strains.** The three O-antigen SNPs investigated differentiated the STEC strains from non-STEC strains, with the exception of seven *stx*-negative strains that displayed the *wzx stx<sub>2</sub>* STEC allele (Table 1). Interestingly, five of these strains also contain *espK*, a virulence-associated effector translocated by the type III secretion system, which substantiates their classification as EHEC-like strains (49). One *stx*-positive strain, belonging to DT 1, was also not identified using the STEC-associated polymorphisms. However, this strain was negative for *espK* and all other virulence-associated markers that have previously been investigated (21).

**SNP genotyping of human clinical O26:H11 strains.** Two human strains, 2009C-3612 and 2009C-3689, were identified to harbor only *stx<sub>2a</sub>* and belonged to MLST ST29. These strains exhibited synonymous SNP genotypes with the HUSEC020 strain (see Table S7 in the supplemental material), belonging to CC-1 (Fig. 2). The O-antigen *stx<sub>2</sub>* STEC-associated allele *wzx* was also observed in these strains. The additional five strains were identified as *stx<sub>1</sub>* and belonging to the MLST ST21. These human clinical strains clustered into CC-4. Four strains, 2009C-4826, 2009C-3996, 2009C-4760, and CFSAN001629, displayed the same SNP genotype profile as that of strains HUSEC017, HUSEC014, and DEC10c. Strain 2009C-4747 differed at three SNP positions from this group and at an additional five SNPs from the O26:H11 reference strain 11368 (Table S7).

## DISCUSSION

**Investigation into genetic characteristics further substantiates similarities between human and bovine O26:H11 strains.** In this study, 51 primer pairs were designed to target 51 previously identified SNPs for next-generation sequencing via a high-throughput targeted amplicon sequencing approach for phylogenetic inference of AEEC O26:H11. The 48 phylogenetically informative SNPs investigated in this study had been observed to form four distinct SNP CCs in EHEC O26 strains and were used to infer an evolutionary model for EHEC O26 (22).

CRISPRs are composed of partially palindromic DNA repeats separated by unique spacers created from exogenous DNA, either from viruses or plasmids, entering the bacterial cell (50, 51). Due to recognition mechanisms of the CRISPR complex, future invasion of this exogenous DNA into the bacterial chromosome is inhibited, therefore creating an adaptive immunity (52). It has also been observed that CRISPR loci are conserved among phylogenetically related bacterial strains (53–55). Previous investigation into the CRISPR 1 and 2a loci of these ST29 AEEC cattle isolates identified that a majority of strains displayed the CRISPR 1 allele 29 and CRISPR 2a allele 67 (21); this CRISPR allele combination had also been previously observed in a *stx<sub>2</sub>*-positive O26 cattle strain (54). The same allele combination has also been identified in the pathogenic human French ST29 *stx<sub>2</sub>* clone displaying the CRISPR PCR marker SP\_O26-E (19). More recently, the same CRISPR allele combination was identified in O26 strains isolated from beef products in processing plants (55), providing evidence that beef products may serve as a route of transmission to humans. The CRISPR 1 and 2a alleles for the U.S. human strains, 2009C-3612 and 2009C-3689, were determined using a CRISPR allele database (21). Strain 2009C-3612 displayed CRISPR 1 allele 29

and CRISPR 2a allele 4, while strain 2009C-3689 displayed CRISPR 1 allele 29 and a CRISPR 2a allele that was not observed in the published database (21). However, this allele, containing the spacer combination 7-11-12-13-14-15, is related to allele 4, spacer combination 7-8-9-10-11-12-13-14-15, given that they share a pool of spacers. The five DT 4 AEEC cattle strains in this study contain the CRISPR type (CT) representing a combination of the CRISPR 1 and 2a alleles, synonymous with that of the human 2009C-3612 strain, CT 4. These cattle strains were also negative for the SP\_O26-E CRISPR PCR marker; this marker was also absent from *stx<sub>2a</sub>* French O26 ST21 and ST29 strains that contained the CRISPR 2a allele 4 (19). Interestingly, the CRISPR 2a allele 4 was previously observed as the most common allele identified in a collection of O26:H11 *stx<sub>1</sub>* strains (54). Investigation of the CRISPR locus identifies similarities and shared characteristics among cattle strains and human clinical strains.

**Expanding the classification of O26:H11 SNP clonal complexes.** Early investigation into the new German EHEC O26 clone (8) identified EHEC O26 ST29 strains carrying only *stx<sub>2a</sub>* to display the plasmid gene profile *ehxA* positive (EHEC hemolysin), *katP* negative (catalase-peroxidase), *espP* negative (serine protease EspP), and *etpD* positive (type II secretion system effector), while the absence of *ehxA* and *etpD* was observed on the plasmid of the new French *stx<sub>2a</sub>* subpopulation (19). With the exception of seven strains (3.95%), all AEEC O26:H11 strains investigated in this study lacked *ehxA*, and 161 strains (90.96%) displayed the characteristic profile of the French clone (21). The clustering of the AEEC strains investigated here in combination with the molecular similarities observed with *stx<sub>2a</sub>* ST29 strains demonstrates that the SNP CCs 1 and 2 may not accurately identify subsets of *stx<sub>2a</sub>* ST29 strains. However, the SNP genotypes and virulence-associated markers present within these cattle AEEC O26:H11 strains and shared with pathogenic human strains, both of U.S. and European origin, provide evidence of the relatedness of these strains. The respective inclusion of three and two cattle O26:H11 strains into CC-3 and CC-4, along with the U.S. human clinical isolates into CC-4 was expected, given that these strains are *stx<sub>1</sub>* positive only and belong to ST21.

**Overcoming shortfalls in SNP discovery studies.** We acknowledge that the SNPs used in this genotyping assay, with the exception of one strain, were identified in strains of European origin, while the cattle strains investigated here belong to a U.S. sample population, and the inclusive and exclusive panel of strains for WGS and SNP discovery is important for identifying markers that are accurate for interpretation of phylogenetic information (24). However, inclusion of the U.S. human clinical isolates, of different Shiga toxin and MLST types, in this study substantiates our observations of the AEEC cattle O26:H11 strains being more closely related to highly pathogenic *stx<sub>2</sub>* human strains than to *stx<sub>1</sub>* strains.

A deeper phylogenetic relationship among and between strains could be achieved with the SNP genotyping assay through addressing limitations in the SNP discovery study. In the initial study, the 48 synapomorphic SNPs were chosen manually from four quarters of the *E. coli* O26:H11 strain 11368 chromosome (22). Some of these SNPs do not provide additional discriminatory information when used in combination with other specific SNPs in this assay and could be substituted for more phylogenetically distinct and informative SNPs. The locus tag combinations containing these less informative SNPs are provided in Table S8 in

the supplemental material. Within these specific locus tag combinations, there is not an observed difference that would result in the discrimination of strains by the inclusion or exclusion of one locus tag with another. However, it is acknowledged that an addition of epidemiologically unrelated strains in this study could result in the necessary inclusion of the SNP positions above. This is supported by the findings presented here for the cattle DT3B strain at the *ycjF* SNP (ECO26\_1890) position in which the observed nucleotide change at this position resulted in a SNP genotype specific to this strain, therefore resulting in a new node to be formed in the phylogenetic minimum spanning tree. Alternatively, rather than excluding these SNPs, the number of primer pairs can be multiplexed in the Access Array assay to increase those investigated. The Access Array system allows up to 10 primer pairs to be multiplexed together in a single primer inlet well. Indeed, the multiplexing capabilities of the Access Array assay can allow the investigation of up to 480 loci per sample, performing 23,040 PCRs on a single IFC.

**Evaluating the relatedness of AEEC O26:H11 cattle and human clinical O26:H11 strains.** As expected, the *stx*<sub>2</sub>-positive U.S. human clinical strains 2009C-3612 and 2009C-3689 exhibit only the *wzx* STEC-associated allele. The seven *stx*-negative O26:H11 cattle strains that displayed the *wzx* STEC-associated allele are in accordance with previous findings in which two strains carrying *espK* were captured (23) and with previous research identifying this gene as a relevant marker for identifying typical EHEC and EHEC-like strains (49). Further, these *stx*-negative EHEC-like strains may have contained or have the potential to acquire a Shiga toxin gene (6, 56, 57). The combination of *espK* and the *wzx* STEC-associated allele further substantiates that these *stx*-negative strains are phylogenetically related to the *stx*-positive strains. The other two *stx*-negative strains that displayed the *wzx* STEC allele belong to DT 12 and contain the SP\_026-E CRISPR PCR marker. While the other two STEC-associated alleles, *rmlA* and *fnl-1*, were positively identified in all other human clinical and *stx*-positive cattle strains, it is important to recognize that the genes containing these discriminative polymorphisms, in addition to O26, are found in other serogroups (23).

**Application of targeted amplicon sequencing for SNP genotyping.** Targeted amplicon sequencing (TAS) provides a pipeline for data interpretation and analysis with limited bioinformatics background when used in combination with a commercially available variant detection program. In combination with NGS technology, TAS increases the throughput of samples, decreases cost, and expedites analysis compared to traditional sequencing methods (58, 59), thus allowing expedited data interpretation and an increased number of strains to be investigated. The application of TAS can be extended beyond the SNPs investigated in this study to include SNPs targeting the emergence of new clones, genomic regions related to physiology, outbreak investigations, and as application in studying antibiotic resistance. We acknowledge that this method is not a replacement for exploratory research aimed at the identification of unique characteristics but serves as a method to increase throughput and identification leading to phylogenetic classification or relatedness to human clinical isolates.

The use of the Access Array in the TAS procedure presented here provides an avenue of immense sample throughput due to its reduced cost benefits accomplished from the small amount of input DNA required, yielding small reagent volumes per sample,

massive multiplexing capabilities, and reduced hands-on time. Furthermore, the primers designed to target the published SNPs investigated in this work can still be used for NGS of these SNPs using other library preparation methods, although they may be more labor-intensive. Alternatively, the primers can also be employed in a high-resolution melting (HRM) assay. This assay utilizes a HRM dye that binds double-stranded DNA (dsDNA) heteroduplexes to detect polymorphisms and generates melting curves of post-PCR products. The presence of different nucleotides affects the melting behavior, leading to the recognition of discernible differences in genotypes. Further, HRM laboratory equipment may currently be more widely available and multipurpose than that previously proposed by the TAS. To corroborate this application, we performed an HRM assay targeting a few SNPs on a subset of strains. A discernible difference was observed in the melting curves produced, indicating a nucleotide change within the targeted region (data not published). Further, the application of HRM has previously been applied in numerous scientific applications (60–64).

**Incorporating an Illumina adaptor-ligated library to increase sequence diversity and read quality.** Less than 1% of the targeted loci failed to be amplified in this study representing no more than one missing locus per strain, except for two strains. However, this missing SNP information had no negative consequences in strain classification for determining phylogenetic relationships of strains within and between clonal complexes. The occurrence of failed locus reads may have been due to the absence of said loci, such as *murC* and *rcsA*, in the genome of the *E. coli* samples. A technical problem due to the lack of sequence diversity could also have contributed to the absence of these results during sequencing, given that conserved sequences that contained very little difference in single nucleotide positions were investigated. To account for this, the Illumina PhiX control, an adaptor-ligated library of the small well-characterized PhiX virus genome, could be incorporated to balance the overall lack of sequence diversity and to increase the quality of sequence reads produced. This can be performed by diluting the denatured 20 pM PhiX library with prechilled HT1 to a final concentration of 12.5 pM. The PhiX control is then combined at 15% (90  $\mu$ l) with the sample library (510  $\mu$ l). Because the PhiX library is read with the Illumina sequencing primers, the custom primers, FL1 and FL2, are prepared as previously described and incorporated with the Illumina primers in the MiSeq cartridge to read the PhiX control during sequencing. It is important to note that these custom primers are not used in the MiSeq custom ports and the Sequencing Workflow Using Fluidigm FL1 and FL2 Sequencing Primers protocol in the *Access Array System for Illumina Sequencing Systems User Guide* (36) should not be followed if using the PhiX control in a sequencing run. Instead, the FL1 and FL2 reagents should be prepared for sequencing on the MiSeq instrument with PhiX according to the protocol described in Table S9 in the supplemental material.

**Conclusion.** Based on the genotypes observed within the 48 phylogenetically informative SNPs, these strains were classified into four previously identified clonal complexes (22). Of these strains, the U.S. cattle AEEC O26:H11 strains displayed SNP genotypes identical to those of *stx*<sub>2</sub> ST29 U.S. human O26:H11 strains. The *stx*<sub>1</sub> ST21 human and cattle strains clustered separately. Although the extent of the clinical outcome of the U.S. human strains is unknown, the similarities of these strains with

the Shiga toxin-negative AEEC cattle strains demonstrate their close phylogenetic relatedness. The potential ability of Shiga toxin-negative *E. coli* strains to acquire a Shiga toxin (57, 65), and the presence of the STEC alleles in a few *stx*-negative cattle strains, further supports the relatedness of AEEC cattle strains to highly pathogenic human strains. NGS has provided an in-depth and robust method for investigation into the bacterial genome. Furthermore, analysis and classification of highly pathogenic bacteria using SNP assays can expedite the time to identification, which is important for food safety and human public health.

## ACKNOWLEDGMENTS

We thank Aubin Fleiss for the use of his local BLAST (BL0Pv2) software program and Josh Ison for helping to construct the R script for variant analysis. We also thank Yannick Marie and Delphine Bouteiller from the iGenSeq Genotyping and Sequencing Core facility at the ICM Brain and Spine Institute for generating the tagged amplicon libraries for sequencing.

This study was supported in part by Agriculture and Food Research Initiative grants (2011-67005-30004 and 2012-68003-30155) from the National Institute of Food and Agriculture, U.S. Department of Agriculture, and by the French Joint ministerial program of R&D against CBRNE risks (grant C17609-2).

We declare that we have no competing financial interests.

## FUNDING INFORMATION

French Joint Ministerial Program of R&D against CBRNE Risks provided funding to Patrick Fach under grant number C17609-2. USDA | National Institute of Food and Agriculture (NIFA) provided funding to Guy H. Loneragan under grant numbers 2011-67005-30004 and 2012-68003-30155.

## REFERENCES

- Brooks JT, Sowers EG, Wells JG, Greene KD, Griffin PM, Hoekstra RM, Strockbine NA. 2005. Non-O157 Shiga-toxin producing *Escherichia coli* infections in the United States, 1983-2002. *J Infect Dis* 192:1422-1429. <http://dx.doi.org/10.1086/466536>.
- European Food Safety Authority-European Centre for Disease Prevention and Control. 2014. The European Union summary report on trends and sources of zoonoses, zoonotic agents and food-borne outbreaks in 2012. *Efsa J* 12:3547.
- Mellmann A, Bielaszewska M, Köch R, Friedrich AW, Fruth A, Middendorf B, Harmsen D, Schmidt MA, Karch H. 2008. Analysis of collection of hemolytic uremic syndrome-associated enterohemorrhagic *Escherichia coli*. *Emerg Infect Dis* 14:1287-1290. <http://dx.doi.org/10.3201/eid1408.071082>.
- Chase-Topping ME, Rosser T, Allison LJ, Courcier E, Evans J, McKendrick JJ, Pearce MC, Handel I, Caprioli A, Karch H, Hanson MF, Pollock KGJ, Locking ME, Woolhouse MEJ, Matthews L, Low JC, Gally DL. 2012. Pathogenic potential to humans of bovine *Escherichia coli* O26, Scotland. *Emerg Infect Dis* 18:439-448. <http://dx.doi.org/10.3201/eid1803.111236>.
- Hussein HS. 2007. Prevalence and pathogenicity of Shiga toxin-producing *Escherichia coli* in beef cattle and their products. *J Anim Sci* 85:E63-E72. <http://dx.doi.org/10.2527/jas.2006-421>.
- Bielaszewska M, Middendorf B, Köck R, Friedrich AW, Fruth A, Karch H, Schmidt MA, Mellmann A. 2008. Shiga toxin-negative attaching and effacing *Escherichia coli*: distinct clinical associations with bacterial phylogeny and virulence traits and inferred in-host pathogen evolution. *Clin Infect Dis* 47:208-217. <http://dx.doi.org/10.1086/589245>.
- Fröhlicher E, Krause G, Zweifel C, Beutin L, Stephan R. 2008. Characterization of attaching and effacing *Escherichia coli* (AEEC) isolated from pigs and sheep. *BMC Microbiol* 8:144-149. <http://dx.doi.org/10.1186/1471-2180-8-144>.
- Bielaszewska M, Mellmann A, Bletz S, Zhang W, Köck R, Kossow A, Prager R, Fruth A, Orth-Höller D, Marejková M, Morabito S, Caprioli A, Piérard D, Smith G, Jenkins C, Čurová K, Karch H. 2013. Enterohemorrhagic *Escherichia coli* O26: H11/H-: a new virulent clone emerges in Europe. *Clin Infect Dis* 56:1373-1381. <http://dx.doi.org/10.1093/cid/cit055>.
- Zhang W-L, Bielaszewska M, Liesegang A, Tschäpe H, Schmidt H, Bitzan M, Karch H. 2000. Molecular characteristics and epidemiological significance of Shiga toxin-producing *Escherichia coli* O26 strains. *J Clin Microbiol* 38:2134-2140.
- Zweifel C, Cernela N, Stephan R. 2013. Detection of the emerging Shiga toxin-producing *Escherichia coli* O26:H11/H- sequence type (ST29) clone in human patients and healthy cattle in Switzerland. *Appl Environ Microbiol* 79:5411-5413. <http://dx.doi.org/10.1128/AEM.01728-13>.
- Sobieszczanska BM, Gryko R, Malek CW. 2000. Isolation of verotoxinogenic strains of *Escherichia coli* O26 in Poland. *Clin Microbiol Infect* 6:227-229. <http://dx.doi.org/10.1046/j.1469-0691.2000.00060-1.x>.
- Paciorek J. 2002. Virulence properties of *Escherichia coli* faecal strains isolated in Poland from healthy children and strains belonging to serogroups O18, O26, O44, O86, O126 and O127 isolated from children with diarrhoea. *J Med Microbiol* 51:548-556. <http://dx.doi.org/10.1099/0022-1317-51-7-548>.
- Allerberger F, Friedrich AW, Grif K, Dierich MP, Dornbusch H-J, Mache CJ, Nachbaur E, Freilinger M, Rieck P, Wagner M, Capriolo A, Karch H, Zimmerhackl LB. 2003. Hemolytic-uremic syndrome associated with enterohemorrhagic *Escherichia coli* O26:H infection and consumption of unpasteurized cow's milk. *Int J Infect Dis* 7:42-45. [http://dx.doi.org/10.1016/S1201-9712\(03\)90041-5](http://dx.doi.org/10.1016/S1201-9712(03)90041-5).
- Ethelberg S, Olsen KEP, Scheutz F, Jensen C, Schiellerup P, Enberg J, Petersen AM, Olesen B, Gerner-Smidt P, Mølbak K. 2004. Virulence factors for hemolytic uremic syndrome, Denmark. *Emerg Infect Dis* 10:842-847. <http://dx.doi.org/10.3201/eid1005.030576>.
- Liptakova A, Siegfried L, Kmetova M, Birosova E, Kotulova D, Benčatova A, Kosecka M, Banovcin P. 2005. Hemolytic uremic syndrome caused by verotoxin-producing *Escherichia coli* O26. Case report. *Folia Microbiol (Praha)* 50:95-98. <http://dx.doi.org/10.1007/BF02931454>.
- Käppli U, Hächler H, Giezendanner N, Beutin L, Stephan R. 2011. Human infections with non-O157 Shiga toxin-producing *Escherichia coli*, Switzerland, 2000-2009. *Emerg Infect Dis* 17:180-185. <http://dx.doi.org/10.3201/eid1702.100909>.
- Verstraete K, De Reu K, Van Weyenberg S, Piérard D, De Zutter L, Herman L, Robyn J, Heyndrickx M. 2013. Genetic characteristics of Shiga toxin-producing *E. coli* O157, O26, O103, O111 and O145 isolates from humans, food, and cattle in Belgium. *Epidemiol Infect* 141:2503-2515. <http://dx.doi.org/10.1017/S0950268813000307>.
- Rivas M, Miliwebsky E, Chinen I, Roldan CD, Balbi L, Garcia B, Fiorilli G, Sosa-Estani S, Kincaid J, Rangel J, Griffin PM. 2006. Characterization and epidemiological subtyping of Shiga toxin-producing *Escherichia coli* strains isolated from hemolytic uremic syndrome and diarrhea cases in Argentina. *Foodborne Pathog Dis* 3:88-96. <http://dx.doi.org/10.1089/fpd.2006.3.88>.
- Delannoy S, Mariani-Kurkdjian P, Bonacorsi S, Liguori S, Fach P. 2015. Characteristics of emerging human-pathogenic *Escherichia coli* O26:H11 isolated in France between 2010 and 2013 and carrying the *stx*<sub>2d</sub> gene only. *J Clin Microbiol* 53:486-492. <http://dx.doi.org/10.1128/JCM.02290-14>.
- Bugarel M, Martin A, Fach P, Beutin L. 2011. Virulence gene profiling of enterohemorrhagic (EHEC) and enteropathogenic (EPEC) *Escherichia coli* strains: a basis for molecular risk assessment of typical and atypical strains. *BMC Microbiol* 11:142-152. <http://dx.doi.org/10.1186/1471-2180-11-142>.
- Ison SA, Delannoy S, Bugarel M, Nightingale KK, Webb HE, Renter DG, Nagaraja TG, Loneragan GH, Fach P. 2015. Genetic diversity and pathogenic potential of attaching and effacing *Escherichia coli* O26:H11 strains recovered from bovine feces in the United States. *Appl Environ Microbiol* 81:3671-3678. <http://dx.doi.org/10.1128/AEM.00397-15>.
- Bletz S, Bielaszewska M, Leopold SR, Koch R, Wittne A, Schuldes J, Zhang W, Karch H, Mellmann A. 2013. Evolution of enterohemorrhagic *Escherichia coli* O26 based on single-nucleotide polymorphisms. *Genome Biol Evol* 5:1807-1816. <http://dx.doi.org/10.1093/gbe/evt136>.
- Norman KN, Strockbine NA, Bono JL. 2012. Association of nucleotide polymorphisms within the O-antigen gene cluster of *Escherichia coli* O26, O45, O103, O111, O121, and O145 with serogroups and genetic subtypes. *Appl Environ Microbiol* 78:6689-6703. <http://dx.doi.org/10.1128/AEM.01259-12>.
- Garvin MR, Saitoh K, Gharrett AJ. 2010. Application of single nucleotide



- polymorphisms to non-model species: a technical review. *Mol Ecol Res* 10:915–934. <http://dx.doi.org/10.1111/j.1755-0998.2010.02891.x>.
25. Seeb JE, Carvalho G, Hauser L, Naish K, Roberts S, Seeb LW. 2011. Single-nucleotide polymorphism (SNP) discovery and applications of SNP genotyping in nonmodel organisms. *Mol Ecol Resour* 11(Suppl 1): 1–8. <http://dx.doi.org/10.1111/j.1755-0998.2010.02979.x>.
  26. Norman KN, Clawson ML, Strockbine N, Mandrell RE, Johnson R, Ziebell K, Zhao S, Fratamico PM, Stones R, Allard MW, Bono JL. 2015. Comparison of whole genome sequences from human and non-human *Escherichia coli* O26 strains. *Front Cell Infect Microbiol* 5:21. <http://dx.doi.org/10.3389/fcimb.2015.00021>.
  27. Hirakawa H, Shirasawa K, Ohyama A, Fukioka H, Aoki K, Rothan C, Sato S, Isobe S, Tabata S. 2013. Genome-wide SNP genotyping to infer the effects on gene functions in tomato. *DNA Res* 20:221–233. <http://dx.doi.org/10.1093/dnares/dst005>.
  28. Stahl PL, Lundeberg J. 2012. Toward the single-hour high-quality genome. *Annu Rev Biochem* 81:359–378. <http://dx.doi.org/10.1146/annurev-biochem-060410-094158>.
  29. Loman NJ, Constantinidou C, Chan JZM, Halachev M, Sergeant M, Penn CW, Robinson ER, Pallen MJ. 2012. High-throughput bacterial genome sequencing: an embarrassment of choice, a world of opportunity. *Nat Rev Microbiol* 10:599–606. <http://dx.doi.org/10.1038/nrmicro2850>.
  30. Franz E, Delaquis P, Morabito S, Beutin L, Gobius K, Rasko DA, Bono J, French N, Osek J, Lindstedt B-A, Muniesa M, Manning S, LeJeune J, Callaway T, Beatson S, Eppinger M, Dallman T, Forbes KJ, Aarts H, Pearl DL, Gannon VP, Laing CR, Strachan NJ. 2014. Exploiting the explosion of information associated with whole genome sequencing to tackle Shiga toxin-producing *Escherichia coli* (STEC) in global food production systems. *Int J Food Microbiol* 187:57–72. <http://dx.doi.org/10.1016/j.ijfoodmicro.2014.07.002>.
  31. Collins FS, Green ED, Guttmacher AE. 2003. A vision for the future of genomics research. *Nature* 422:835–847. <http://dx.doi.org/10.1038/nature01626>.
  32. Ju W, Cao G, Rump L, Strain E, Luo Y, Timme R, Allard M, Zhao S, Brown E, Meng J. 2012. Phylogenetic analysis on non-O157 Shiga toxin-producing *Escherichia coli* strains by whole-genome sequencing. *J Clin Microbiol* 50:4123–4127. <http://dx.doi.org/10.1128/JCM.02262-12>.
  33. Paddock ZD, Renter DG, Cull CA, Shi X, Bai J, Nagaraja TG. 2014. *Escherichia coli* O26 in feedlot cattle: fecal prevalence, isolation, characterization, and effects of an *E. coli* O157 vaccine and a direct-fed microbial. *Foodborne Pathog Dis* 11:186–193. <http://dx.doi.org/10.1089/fpd.2013.1659>.
  34. Ogura Y, Ooka T, Iguchi A, Toh H, Asadulghani M, Oshima K, Kodama T, Abe H, Nakayama K, Kurokawa K, Tobe T, Hattori M, Hayashi T. 2009. Comparative genomics reveal the mechanism of the parallel evolution of O157 and non-O157 enterohemorrhagic *Escherichia coli*. *Proc Natl Acad Sci U S A* 106:17939–17944. <http://dx.doi.org/10.1073/pnas.0903585106>.
  35. Fluidigm Corporation. 2015. Access Array Superior library preparation on the widest scale with a streamlined workflow and low costs. Fluidigm, South San Francisco, CA. <https://www.fluidigm.com/products/access-array>. Accessed 3 July 2015.
  36. Fluidigm Corporation. 2015. Access Array system for Illumina sequencing systems user guide. PN 100-3770 H1. Fluidigm Corporation, South San Francisco, CA.
  37. Untergasser A, Nijveen H, Rao X, Bisseling T, Geurts R, Leunissen JAM. 2007. Primer3Plus, an enhanced web interface to Primer3. *Nucleic Acids Res* 35:W71–W74. <http://dx.doi.org/10.1093/nar/gkm306>.
  38. Kibbe WA. 2007. OligoCalc: an online oligonucleotide properties calculator. *Nucleic Acids Res* 35:W43–W6. <http://dx.doi.org/10.1093/nar/gkm234>.
  39. Fluidigm Corporation. 2015. Fluidigm Access Array - generate tagged primers workbook. PN 100-3873. Fluidigm Corporation, South San Francisco, CA.
  40. Fluidigm Corporation. 2015. Appendix C. 2-Primer target-specific PCR amplification, p 85–100. Access Array system for Illumina sequencing systems user guide. PN 100-3770. Fluidigm Corporation, South San Francisco, CA.
  41. Illumina, Inc. 2013. Preparing libraries for sequencing on the MiSeq. Part 150369740. Revision D. Illumina, Inc., San Diego, CA.
  42. Illumina, Inc. 2012. Illumina MiSeq user guide. Part 15027617. Illumina, Inc., San Diego, CA.
  43. Huson DH, Bryant D. 2006. Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol* 23:254–267.
  44. Trees E, Strockbine N, Changayil S, Ranganathan S, Zhao K, Weil R, MacCannell D, Sabol A, Schmidtke A, Martin H, Stripling D, Ribot EM, Gerner-Smidt P. 2014. Genome sequences of 228 Shiga toxin-producing *Escherichia coli* isolates and 12 isolates representing other diarrheagenic *E. coli* pathotypes. *Genome Announc* 2:e00718-14. <http://dx.doi.org/10.1128/genomeA.00718-14>.
  45. Cao G, Ju W, Rump L, Zhao S, Zou L, Wang C, Strain E, Luo Y, Timme R, Allard M, Brown E, Meng J. 2013. Genome sequences of two emerging non-O157 Shiga toxin-producing *Escherichia coli* strains. *Genome Announc* 1:e00200-13. <http://dx.doi.org/10.1128/genomeA.00200-13>.
  46. Wattam AR, Abraham D, Dalay O, Disz TL, Driscoll T, Gabbard JL, Gillespie JJ, Gough R, Hix D, Kenyon R, Machi D, Mao C, Nordberg EK, Olson R, Overbeek RA, Pusch GD, Shukla M, Schulman J, Stevens RL, Sullivan DE, Vonstein V, Warren A, Will R, Wilson MJ, Yoo HS, Zhang C, Zhang Y, Sobral BW. 2014. PATRIC, the bacterial bioinformatics database and analysis resource. *Nucleic Acids Res* 42:D581–D591. <http://dx.doi.org/10.1093/nar/gkt1099>.
  47. Larsen MV, Cosentino S, Rasmussen S, Friis C, Hasman H, Marvig RL, Jelsbak L, Sicheritz-Pontén T, Ussery DW, Aarestrup FM, Lund O. 2012. Multilocus sequence typing of total-genome-sequenced bacteria. *J Clin Microbiol* 50:1355–1361. <http://dx.doi.org/10.1128/JCM.06094-11>.
  48. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol* 215:401–410.
  49. Bugarel M, Beutin L, Scheutz F, Loukiadis E, Fach P. 2011. Identification of genetic markers for differentiation of Shiga toxin-producing, enteropathogenic, and avirulent strains of *Escherichia coli* O26. *Appl Environ Microbiol* 77:2275–2281. <http://dx.doi.org/10.1128/AEM.02832-10>.
  50. Barrangou R. 2013. CRISPR-Cas systems and RNA-guided interference. *Wiley Interdiscip Rev RNA* 4:267–278. <http://dx.doi.org/10.1002/wrna.1159>.
  51. Bolotin A, Quinquis B, Sorokin A, Ehrlich SD. 2005. Clustered regularly interspaced short palindromic repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology* 151:2551–2561.
  52. Horvath P, Barrangou R. 2010. CRISPR/Cas, the immune system of bacteria and archaea. *Science* 327:167–170. <http://dx.doi.org/10.1126/science.1179555>.
  53. Touchon M, Charpentier S, Clermont O, Rocha EPC, Denamur E, Branger C. 2011. CRISPR distribution within the *Escherichia coli* species is not suggestive of immunity-associated diversifying selection. *J Bacteriol* 193:2460–2467. <http://dx.doi.org/10.1128/JB.01307-10>.
  54. Yin S, Jensen MA, Bai J, DeRoy C, Barrangou R, Dudley EG. 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. <http://dx.doi.org/10.1128/AEM.00950-13>.
  55. Jiang Y, Yin S, Dudley EG, Cutter CN. 2015. Diversity of CRISPR loci and virulence genes in pathogenic *Escherichia coli* isolates from various sources. *Int J Food Microbiol* 204:41–46. <http://dx.doi.org/10.1016/j.ijfoodmicro.2015.03.025>.
  56. Bielaszewska M, Dobrindt U, Gärtner J, Gallitz I, Hacker J, Karch H, Müller D, Schubert S, Schmidt MA, Sorsa LJ, Zdzinski J. 2007. Aspects of genome plasticity in pathogenic *Escherichia coli*. *Int J Med Microbiol* 297:625–639. <http://dx.doi.org/10.1016/j.ijmm.2007.03.001>.
  57. Friedrich AW, Zhang W, Bielaszewska M, Köck R, Fruth A, Tschäpe H, Karch H. 2007. Prevalence, virulence profiles, and clinical significance of Shiga toxin-negative variants of enterohemorrhagic *Escherichia coli* O157 infection in humans. *Clin Infect Dis* 45:39–45. <http://dx.doi.org/10.1086/518573>.
  58. Campbell NR, Harmon SA, Narum SR. 2015. Genotyping-in-thousands by sequencing (GT-seq): a cost effective SNP genotyping method based on custom amplicon sequencing. *Mol Ecol Resour* 15:855–867. <http://dx.doi.org/10.1111/1755-0998.12357>.
  59. Lange V, Böhme I, Hofmann J, Lang K, Sauter J, Schöne B, Paul P, Albrecht V, Andreas JM, Baier DM, Nething J, Ehninger U, Schwarzelt C, Pingel J, Ehninger G, Schmidt AH. 2014. Cost-efficient high-throughput HLA typing by MiSeq amplicon sequencing. *BMC Genomics* 15:63. <http://dx.doi.org/10.1186/1471-2164-15-63>.
  60. Derzelle S. 2015. Single-nucleotide polymorphism discrimination using high-resolution melting analysis for the genotyping of *Bacillus anthracis*. *Methods Mol Biol* 1247:361–371. [http://dx.doi.org/10.1007/978-1-4939-2004-4\\_26](http://dx.doi.org/10.1007/978-1-4939-2004-4_26).
  61. Ilie DE, Cean A, Gavriliuc OI, Carstea AC, Grădinaru AC. 2013. High-resolution melting assay as a tool for identification of CSN3 genotypes in cattle population. *Anim Sci Biotechnol* 46:102–105.
  62. Liew M, Pryor R, Palais R, Meadows C, Erali M, Lyon E, Wittwer C.

2004. Genotyping of single-nucleotide polymorphisms by high-resolution melting of small amplicons. *Clin Chem* 50:1156–1164. <http://dx.doi.org/10.1373/clinchem.2004.032136>.
63. Mazi W, Sangal V, Saeed A, Sandstorm G, Weill F-X, Yu J. 2015. Rapid genotyping of *Shigella sonnei* by use of multiplex high-resolution melting. *J Clin Microbiol* 53:2389–2391. <http://dx.doi.org/10.1128/JCM.00874-15>.
64. Pholwat S, Liu J, Stroup S, Gratz J, Banu S, Rahman SMM, Ferdous SS, Foongladda S, Boonlert D, Ogarkov O, Zhdanova S, Kibiki G, Heysell S, Hout E. 2015. Integrated microfluidic card with TaqMan probes and high-resolution melt analysis to detect tuberculosis drug resistance mutations across 10 genes. *mBio* 6:e02273-14. <http://dx.doi.org/10.1128/mBio.02273-14>.
65. Bielaszewska M, Prager R, Köck R, Mellmann A, Zhang W, Tschäpe H, Tarr PI, Karch H. 2007. Shiga toxin gene loss and transfer in vitro and in vivo during enterohemorrhagic *Escherichia coli* O26 infection in humans. *Appl Environ Microbiol* 73:3144–3150. <http://dx.doi.org/10.1128/AEM.02937-06>.