

Phylogeny of Shiga Toxin-Producing *Escherichia coli* O157 Isolated from Cattle and Clinically Ill Humans

James L. Bono,¹ Timothy P.L. Smith,¹ James E. Keen,² Gregory P. Harhay,¹ Tara G. McDanel,¹ Robert E. Mandrell,³ Woo Kyung Jung,⁴ Thomas E. Besser,⁴ Peter Gerner-Smidt,⁵ Martina Bielaszewska,⁶ Helge Karch,⁶ and Michael L. Clawson^{*1}

¹United States Department of Agriculture, Agricultural Research Service, US Meat Animal Research Center, Clay Center, Nebraska

²Great Plains Veterinary Educational Center, University of Nebraska

³United States Department of Agriculture, Agricultural Research Service, Western Regional Research Center, Albany, California

⁴Department of Veterinary Microbiology and Pathology, Washington State University

⁵Enteric Diseases Laboratory Response Branch, Centers for Disease Control and Prevention, Atlanta, Georgia

⁶Institute of Hygiene and the National Consulting Laboratory on Haemolytic Uraemic Syndrome, University of Münster, Münster, Germany

*Corresponding author: E-mail: mike.clawson@ars.usda.gov.

Associate editor: Helen Piontkivska

Abstract

Cattle are a major reservoir for Shiga toxin-producing *Escherichia coli* O157 (STEC O157) and harbor multiple genetic subtypes that do not all associate with human disease. STEC O157 evolved from an *E. coli* O55:H7 progenitor; however, a lack of genome sequence has hindered investigations on the divergence of human- and/or cattle-associated subtypes. Our goals were to 1) identify nucleotide polymorphisms for STEC O157 genetic subtype detection, 2) determine the phylogeny of STEC O157 genetic subtypes using polymorphism-derived genotypes and a phage insertion typing system, and 3) compare polymorphism-derived genotypes identified in this study with pulsed field gel electrophoresis (PFGE), the current gold standard for evaluating STEC O157 diversity. Using 762 nucleotide polymorphisms that were originally identified through whole-genome sequencing of 189 STEC O157 human- and cattle-isolated strains, we genotyped a collection of 426 STEC O157 strains. Concatenated polymorphism alleles defined 175 genotypes that were tagged by a minimal set of 138 polymorphisms. Eight major lineages of STEC O157 were identified, of which cattle are a reservoir for seven. Two lineages regularly harbored by cattle accounted for the majority of human disease in this study, whereas another was rarely represented in humans and may have evolved toward reduced human virulence. Notably, cattle are not a known reservoir for *E. coli* O55:H7 or STEC O157:H⁻ (the first lineage to diverge within the STEC O157 serogroup), which both cause human disease. This result calls into question how cattle may have originally acquired STEC O157. The polymorphism-derived genotypes identified in this study did not surpass PFGE diversity assessed by *BlnI* and *XbaI* digestions in a subset of 93 strains. However, our results show that they are highly effective in assessing the evolutionary relatedness of epidemiologically unrelated STEC O157 genetic subtypes, including those associated with the cattle reservoir and human disease.

Key words: STEC O157, phylogeny, evolution, disease, cattle, human.

Introduction

Shiga toxin-producing *Escherichia coli* O157 (STEC O157) are genetically diverse bacteria that cause diarrhea, hemorrhagic colitis, and hemolytic uremic syndrome (HUS) in humans (Griffin and Tauxe 1991; Gunzer et al. 1992). STEC O157:H7 and STEC O157:H⁻ strains comprise the STEC O157 serogroup. STEC O157:H7 strains are typically motile, do not ferment sorbitol (SOR⁻), and most do not express β -glucuronidase activity (GUD⁻), whereas STEC O157:H⁻ strains are typically nonmotile, SOR⁺, and GUD⁺ (fig. 1). Both subgroups are recently emerged pathogens; however, STEC O157:H⁻ human infections have been primarily limited to Europe (Alpers et al. 2009; Pollock et al. 2010). In contrast, STEC O157:H7 is responsible for a large majority of human STEC O157 infections throughout the world (Griffin and

Tauxe 1991; Mead and Griffin 1998). Thus, the STEC O157:H7 subgroup, in particular, is an international concern.

Cattle are a major reservoir for STEC O157:H7 and a source of human infection (Borczyk et al. 1987; Griffin and Tauxe 1991; Wells et al. 1991). However, cattle also harbor certain STEC O157:H7 genetic subtypes that are rarely found in clinically ill humans (Kim et al. 1999; Roldgaard et al. 2004; Besser et al. 2007; Bono et al. 2007; Clawson et al. 2009; Whitworth et al. 2010). Cattle occasionally harbor SOR⁺ STEC O157:H⁻, although the primary reservoir for this serotype is unknown (Bielaszewska et al. 2000; Lee and Choi 2006; Alpers et al. 2009). Consequently, cattle and human clinical STEC O157 cases represent important focal points for identifying STEC O157 genetic subtypes, understanding STEC O157 evolution and the evolution of genetic

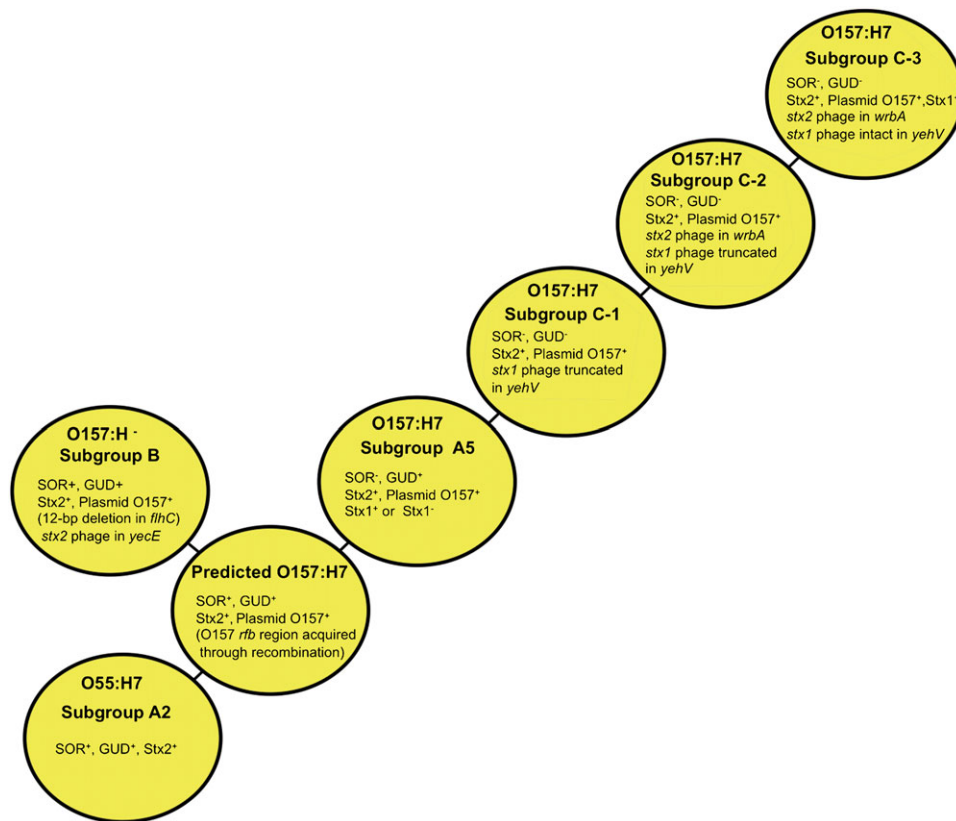


Fig. 1. Model of STEC O157 descent from an STEC O55:H7 progenitor based on previous studies. STEC O157 subgroups within the model are defined by the presence (+) or absence (–) of the ability to ferment SOR, β -glucuronidase activity (GUD), Shiga toxin 1 (Stx1), Shiga toxin 2 (Stx2), and pO157; Stx1-encoding temperate bacteriophage in *yehV* (truncated or complete), Stx2-encoding temperate bacteriophage in *wrbA*, and Stx2-encoding temperate bacteriophage in *yecE*. Subgroups A2, B, and A5 are described by Feng et al. (2007). Subgroups C-1, C-2, and C-3 are described by Shaikh et al. (2007). This model is not to evolutionary scale.

subtypes associated with human disease, and developing an STEC O157 typing system for use in STEC O157 outbreak investigations.

A series of signature events point to the origin of STEC O157 and subsequent evolution within the serogroup (fig. 1). STEC O157 is thought to have evolved from a SOR⁺, GUD⁺ O55:H7 predecessor that acquired a bacteriophage resulting in Shiga toxin 2 expression (Stx2⁺) (fig. 1, subgroup A2; Feng et al. 2007). The O55:H7 predecessor is thought to have given rise to an unobserved O157:H7 lineage predicted to be Stx2⁺, SOR⁺, and GUD⁺ that acquired an *rfb* gene cluster that encoded somatic O antigens of the O157 serogroup through recombination and was transformed with plasmid pO157 (92–104 kb), (fig. 1, predicted O157:H7; Tarr et al. 2000; Shaikh and Tarr 2003; Feng et al. 2007; Leopold et al. 2009; Lim et al. 2010). The predicted lineage is thought to have given rise to two extant STEC O157 lineages: 1) nonmotile O157:H[–] that are Stx2⁺, SOR⁺, GUD⁺, pO157⁺, contain a 12-bp deletion in the flagellar regulatory gene (*flhC*), and have a temperate *stx2*-containing bacteriophage inserted within *yecE*, a gene with unknown biological function (fig. 1, subgroup B; Shaikh and Tarr 2003; Bielaszewska et al. 2006; Feng et al. 2007; Leopold et al. 2009) and 2) O157:H7 that are Stx2⁺, SOR[–], GUD[–], pO157⁺, and may be either positive or negative for Shiga toxin 1 (Stx1) expression (fig. 1, subgroup A5; Monday et al.

2001; Feng et al. 2007). Three extant lineages are thought to have sequentially evolved from subgroup A5. Subgroup C1 members are Stx2⁺, SOR[–], GUD[–], pO157⁺, and contain truncated temperate bacteriophages that are inserted into *yehV*, which encodes a transcriptional regulator for mannose-resistant hemagglutination and curli production (Brown et al. 2001; Shaikh et al. 2007; Leopold et al. 2009). Subgroup C2 members differ from subgroup C1 by having a temperate *stx2*-containing phage integrated within *wrbA*, which encodes a multimeric flavodoxin-like protein (Grandori et al. 1998; Shaikh et al. 2007; Leopold et al. 2009). Subgroup C3 members differ from subgroup C2 by being Stx1⁺, with intact *stx1*-containing temperate phage inserted into *yehV* (Shaikh et al. 2007; Leopold et al. 2009). These events, although informative, define STEC O157 evolution at a relatively low resolution in relation to the number of nucleotide polymorphisms that are known within the STEC O157 serogroup (Zhang et al. 2006; Bono et al. 2007; Jackson et al. 2007; Clawson et al. 2009; Leopold et al. 2009). Additionally, the evolution of an entire lineage of STEC O157 genetic subtypes that is maintained in cattle and does not associate with human disease has not been resolved using these signature events (Bono et al. 2007; Clawson et al. 2009).

Nucleotide polymorphism–derived genotyping shows great potential in distinguishing STEC O157 genetic subtypes,

including those that associate with either the cattle reservoir and/or increased human virulence, and in determining the course of evolution within the serogroup (Manning et al. 2008; Clawson et al. 2009; Leopold et al. 2009). To date, a clade of genetic subtypes defined by a set of 96 nucleotide polymorphisms has been associated with a higher proportion of HUS cases in humans (Manning et al. 2008). Additionally, a different clade defined by a set of 178 polymorphisms has been identified that does not associate with human disease (Clawson et al. 2009). Eventually, nucleotide polymorphism-derived genotyping may replace pulsed field gel electrophoresis (PFGE) as a standard method for detecting STEC O157 genetic subtypes in outbreak investigations. Unlike PFGE, which is most useful in analyzing epidemiologically related strains (Davis, Hancock, Besser, and Call 2003), nucleotide polymorphism-derived genotyping can be used to determine genetic relatedness between strains whether they are epidemiologically related or not (Clawson et al. 2009; Leopold et al. 2009). However, greater numbers of informative nucleotide polymorphisms are needed to resolve diverse STEC O157 genetic subtypes and to map evolution within the serogroup (Clawson et al. 2009).

The main goals of this study were to identify nucleotide polymorphisms for high-resolution evolutionary analyses of STEC O157 genetic subtypes typically found in either cattle or clinically ill humans and to compare polymorphism-derived genotype diversity directly with PFGE diversity. Here, we report 584 newly described polymorphisms that were genotyped in combination with 178 previously characterized polymorphisms (762 total polymorphisms) in 426 STEC O157 strains ($n = 269$ cattle, 154 human, 1 pig, 1 soil, and 1 water). One hundred and seventy-five polymorphism-derived genotypes were observed that are detectable by a minimal set of 138 polymorphisms. Using phylogenetic analyses of the genotypes, we identified eight major STEC O157 lineages. Seven lineages are typically found in cattle, including one that does not associate with human disease and may be evolving away from human virulence and two others that account for a majority of human disease. PFGE diversity surpassed that from the polymorphism-derived genotypes identified in this study. However, these results highlight the tremendous utility of nucleotide polymorphism-based genotyping as a method to determine evolutionary relatedness within the STEC O157 serogroup.

Materials and Methods

Bacterial Strains

A total of 426 STEC O157 strains ($n = 269$ isolated from cattle, 154 isolated from humans, 1 isolated from pig, 1 isolated from soil, and 1 isolated from water) were used in this study for nucleotide polymorphism genotyping and phylogenetic analyses (supplementary table 1, Supplementary Material online) (Whittam 1998; Reid et al. 1999; Elder et al. 2000; Hayashi et al. 2001; Perna et al. 2001; Iguchi et al. 2002; Davis, Hancock, Besser, Rice, et al. 2003;

Bono et al. 2004; Keen et al. 2006; Cooley et al. 2007; Clawson et al. 2009). Moreover, 96 of the 426 strains were used for PFGE analyses and 96 of the 426 were genotyped for Stx-encoding bacteriophage insertion (SBI) sites (supplementary table 1, Supplementary Material online). The 96 strains used for PFGE analyses originated from routine United States Department of Agriculture surveys of slaughter plants within the United States. All 426 strains were verified as STEC O157 via an enzyme-linked immunosorbent assay using an anti-O157 monoclonal antibody and multiplex polymerase chain reaction (PCR) for *stx1*, *stx2*, the intimin gene (*eae*), the plasmid-encoded enterohemolysin gene of enterohemorrhagic *E. coli* (EHEC-*hlyA*), *rfb*_{O157}, and the flagellin gene (*fliC*_{H7}) (He et al. 1996; Gannon et al. 1997; Westerman et al. 1997; Paton AW and Paton JC 1998). Additionally, each strain was genotyped for a translocated intimin receptor gene (*tir* 255T > A) polymorphism (Bono et al. 2007). One STEC O55:H7 strain was also used for single-nucleotide polymorphism genotyping. This strain (815-02) originates from Germany and was PCR positive for *stx2*, *eae*, and *fliC*_{H7}; PCR negative for *stx1*, EHEC-*hlyA*, and *rfb*_{O157} (Gannon et al. 1997; Paton AW and Paton JC 1998); and O55 positive via *rfb* sequence analysis.

Nucleotide Polymorphism Discovery and Selection

All 762 nucleotide polymorphisms genotyped in this study were originally identified through 1X whole-genome sequencing coverage of 189 STEC O157 strains, of which 188 were SOR⁻ STEC O157:H7 (Clawson et al. 2009) (supplementary table 1, Supplementary Material online). These 189 strains are part of the larger set of 426 strains used in this study. Of the 189 strains, 102 originate from cattle that were distributed throughout different locations within the United States. The remaining 87 strains all originate from clinically ill humans and were selected based on PFGE patterns and/or epidemiological data. Some strains within the set of 87 are epidemiologically linked and were excluded from frequency estimates of nucleotide polymorphism alleles, polymorphism-derived genotypes, and SBI genotypes in humans (supplementary table 1, Supplementary Material online). The STEC O157 genomes were sequenced in one of three DNA pools that accounted for host source (cattle or human) and genotype for the *tir* 255T > A polymorphism. The A allele of this polymorphism is an effective marker for a lineage of STEC O157 strains that is maintained in cattle and rarely observed in clinically ill humans, whereas the T allele marks several lineages that are found in both clinically ill humans and cattle (Bono et al. 2007; Clawson et al. 2009). Genomic libraries were prepared for each of the three pools in accordance with the manufacturer's protocols (Roche, Nutley, NJ), and the sequencing was performed on the Roche GS-FLX platform. The first pool contained DNAs from 51 cattle-isolated STEC O157:H7 strains with the A allele of *tir* 255 T > A and produced 346.2 Mb of sequence. The second pool contained DNAs from 51 cattle-isolated STEC O157:H7 strains with the T allele of *tir* 255 T > A and produced 402.6 Mb of

sequence. The third pool contained DNAs from 86 STEC O157:H7 strains with the T allele of *tir* 255T > A that were isolated from clinically ill humans and one human-isolated SOR⁺ STEC O157:H⁻ strain with the T allele of *tir* 255T > A and produced 557.5 Mb of sequence (Clawson et al. 2009). This design allowed for the identification of STEC O157 genetic diversity that was from cattle and not typically found in clinically ill humans, from cattle and may or may not be found in clinically ill humans, and from clinically ill humans. A total of 16,218 putative nucleotide and/or insertion deletion polymorphisms were identified across the pools by mapping individual reads onto the Sakai reference strain (GenBank: NC_002695) with Roche GS Reference Mapper Software (version 1.1.03, Roche) (Clawson et al. 2009).

Selection and validation of 762 nucleotide polymorphisms from 16,218 putative candidates were based on genome location and minor allele frequencies within or across the DNA pools and was conducted over this study and a previous one (Clawson et al. 2009). In this study, we genotyped and validated 584 of the putative polymorphisms. We also genotyped an additional 178 polymorphisms that were previously validated from the list of 16,218 putative polymorphisms (Clawson et al. 2009). Regarding the overall selection of 762 nucleotide polymorphisms, we primarily targeted those that reside on a 4.1 Mb backbone of STEC O157 that is shared with *E. coli* K-12 and other *E. coli* serogroups ($n = 616$) versus those present on S-loops, which are segments of the STEC O157 genome that are absent in the *E. coli* K-12 genome (Hayashi et al. 2001; Perna et al. 2001) ($n = 146$) (supplementary table 2, Supplementary Material online). This allowed us to identify polymorphisms that would have informativity that extended beyond the STEC O157 serogroup. To capture STEC O157 genetic diversity within each of the DNA pools, 398 nucleotide polymorphisms were selected where the minor allele was observed exclusively in one pool. Of these, 77 were selected from the *tir* 255T > A T human strain DNA pool, 161 from the *tir* 255T > A T bovine strain pool and 160 from the *tir* 255T > A A bovine strain pool (fig. 2). Additionally, 340 nucleotide polymorphisms were selected that shared minor alleles between the human strain DNA pool and the bovine strain *tir* 255T > A T DNA pool, and an additional six were selected with shared minor alleles between the two bovine DNA pools (fig. 2). Seventeen selected nucleotide polymorphisms had minor alleles that were observed across all three pools (fig. 2).

Nucleotide Polymorphism Genotyping

STEC O157 strains ($n = 426$) were genotyped for 762 polymorphism alleles either by sequencing on an ABI 3730 capillary sequencer (PE Applied Biosystems, Foster City, CA) or by matrix-assisted laser desorption ionization–time-of-flight (MALDI-TOF) genotyping (Sequenom, Inc., San Diego, CA). MALDI-TOF assay and multiplexing design were conducted with MassARRAY assay design software as recommended by the manufacturer (Sequenom, Inc.). Up to 36 polymorphisms were accepted for each mul-

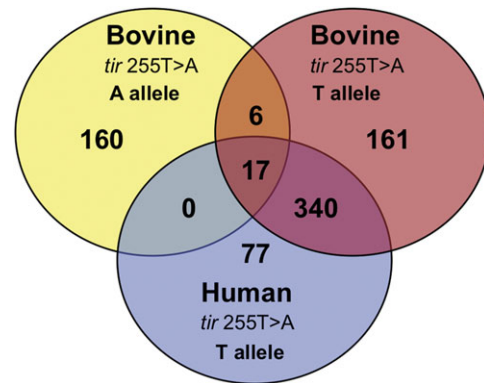


Fig. 2. Number of nucleotide polymorphisms validated from the three STEC O157 DNA pools. The numbers represent nucleotide polymorphisms with minor allele frequencies greater than zero in the DNA pools used for their discovery. One of the 762 nucleotide polymorphisms characterized in this study is not represented in the figure as the minor allele was only observed in the Sakai reference genome (GenBank: NC_002695). Strains used for polymorphism discoveries and validations are listed in supplementary table 1, Supplementary Material online. Nucleotide polymorphisms characterized in this study and their frequencies in human and cattle strains are listed in supplementary table 2, Supplementary Material online.

tiplex, and the assays were conducted with iPLEX Gold chemistry on a MassARRAY genotyping system as per the instructions of the manufacturer (Sequenom Inc.). “High confidence” genotype calls by the Genotyper software were accepted as correct. “Aggressive” calls were inspected manually and verified as needed by replicate MALDI-TOF assays or Sanger Sequencing. Previously designed MALDI-TOF assays were used to genotype 178 of the 762 polymorphisms (Clawson et al. 2009). Additional MALDI-TOF assays were developed in this study to score 584 nucleotide polymorphisms. For each of the 426 STEC O157 strains, all 762 polymorphism alleles were concatenated to generate a polymorphism-derived genotype.

Identification of Haplotype Tagging Alleles

The alleles of 762 nucleotide polymorphisms were concatenated in order against the Sakai STEC O157 genome (GenBank: NC_002685) for 426 strains and aligned using ClustalX (version 1.83) (Thompson et al. 1997). Unique polymorphism-derived genotypes within the alignment ($n = 175$) were identified in Tree-Puzzle (version 5.2) (Schmidt et al. 2002) and analyzed for haplotype tagging alleles in Haploview (version 4.2) (Barrett et al. 2005). Putative polymorphism combinations that tagged all 175 unique polymorphism-derived genotypes were identified with the “Tagger” feature of Haploview using options to capture all alleles through pairwise tagging with $r^2 = 1$ and to accept 32 previously described tagging polymorphisms (Clawson et al. 2009). Using only the putative tagging polymorphisms identified by Haploview, a new alignment of all 175 unique polymorphism-derived genotypes (based on 762 nucleotide polymorphisms) was constructed in ClustalX. To test the actual tagging power of each polymorphism allele, polymorphism alleles were

sequentially subtracted from the alignment and Tree-Puzzle was used to search for redundant sequences. If the subtraction of a nucleotide polymorphism led to the identification of redundant sequences in Tree-Puzzle, the nucleotide polymorphism was accepted as tagging and added back to the alignment. If the subtraction still yielded unique sequences, the polymorphism was not truly tagging and the polymorphism was excluded from further consideration as a tagging polymorphism.

Phylogenetic Analyses

Neighbor joining trees were produced from ClustalX alignments of 175 distinct full-length STEC O157 polymorphism-derived genotypes and one homologous full-length STEC O55:H7 polymorphism-derived genotype ($n = 762$ alleles, [supplementary table 3, Supplementary Material online](#)) and from haplotype tagging polymorphism-derived genotypes for the same strains ($n = 138$ alleles), ([supplementary table 4, Supplementary Material online](#)). Rooted trees were constructed for both alignments in PHYLIP (version 3.69) (Felsenstein J, unpublished data) using the programs CONSENSE, DNADIST, NEIGHBOR, and SEQBOOT. A distance matrix was produced in DNADIST using an F84 distance model of substitution and a transition/transversion ratio of 2. Neighbor joining trees were constructed in NEIGHBOR and rooted with the O55:H7 sequence. Bootstrap values were determined for the trees by generating 1,000 pseudoalignments in SEQBOOT, 1,000 distance matrices in DNADIST, and 1,000 trees in NEIGHBOR. A consensus tree was generated in CONSENSE. Both consensus and individual trees were viewed with Dendroscope (version 2.5) (Huson et al. 2007).

Pulsed-Field Gel Electrophoresis

Standardized protocols for STEC O157 PFGE (Ribot et al. 2006) were applied to 96 STEC O157 strains isolated from ground beef ([supplementary table 1, Supplementary Material online](#)). Two independent digestions were conducted for each strain: a primary with *Xba*I and a secondary with *Bln*I (Ribot et al. 2006). PFGE patterns were analyzed semiautomatically with Bionumerics software (v.5.0; Applied Maths, Austin, TX) and compared with patterns in the PulseNet USA database. Matching PFGE names correspond to indistinguishable PFGE patterns. PFGE scores and polymorphism-derived genotype scores for the 96 strains were independently determined.

SBI Genotyping

Most of the SBI genotypes analyzed in this study were available from the literature, as were their frequencies in STEC O157 strains of human and/or cattle sources (Besser et al. 2007; Whitworth et al. 2008) ([supplementary table 1, Supplementary Material online](#)). Briefly, uniplex PCR assays were conducted to detect the presence or absence of *stx1* and *stx2* genes (Olsvik et al. 1991; Paton AW and Paton JC 1998; Whitworth et al. 2008) and phage integration, full, partial, or lack thereof into either *yehV* (Stx1-encoding bac-

teriophage) or *wrbA* (Stx2-encoding bacteriophage) (Shaikh and Tarr 2003; Besser et al. 2007) ([supplementary table 5, Supplementary Material online](#)). The presence or absence of PCR amplicons was determined via gel electrophoresis. Additionally, for 96 strains SBI genotyped in this study, the uniplex PCRs described above were combined into a single multiplex PCR ([supplementary table 5, Supplementary Material online](#)). Multiplex 25- μ l PCRs consisted of 1.25 units Platinum Taq polymerase (Invitrogen, Carlsbad, CA), 2.4 mM MgCl₂, 0.4 mM deoxynucleoside triphosphates, 2.5 μ l 10 \times buffer (Invitrogen), 0.2 μ M of primers A, E, F, B, C, G, H, and D, 0.3 μ M of primers SLT1-2 and SLT1-1, 0.1 μ M of primers Stx2F and Stx2R ([supplementary table 5, Supplementary Material online](#)), and 5 ng of DNA template. The parameters of the thermocycler (iCycler; Bio-Rad, Hercules, CA) included one 95 $^{\circ}$ C (5 min) cycle and 35 cycles at 94 $^{\circ}$ C (30 s), 58 $^{\circ}$ C (45 s), and 72 $^{\circ}$ C (90 s), followed by a final 72 $^{\circ}$ C (10 min) cycle. After completion of the PCR, samples containing 2 μ l PCR product, 0.5 μ l GeneScan 1200Liz size standard, and 12.5 μ l of Hi-Di formamide were analyzed by capillary electrophoresis using an ABI-3730 DNA Analyzer (PE Applied Biosystems, Foster City, CA) at the Laboratory for Biotechnology and Bioanalysis, Washington State University, Pullman, WA. Following data analysis with GeneMarker 1.70 software (SoftGenetics, State College, PA), genotypes were inferred based on the presence or absence of peaks consistent with each of the six PCR products (Besser et al. 2007; Whitworth et al. 2008). Resultant PCR products were coded for six characters (0 for absence and 1 for presence) for the following concatenated genotypes *stx1*, *stx2*, *yehV*-left and Stx1-encoding bacteriophage junction, *yehV*-right and Stx1-encoding bacteriophage junction, *wrbA*-left and Stx2-encoding bacteriophage junction, and *wrbA*-right and Stx2-encoding bacteriophage junction ([supplementary table 1, Supplementary Material online](#)).

Results

Polymorphism Validation and Genotyping

A total of 762 nucleotide polymorphisms were genotyped across 426 STEC O157 strains. Of the 762, the minor alleles of 183 were observed within the bovine strain *tir* 255T > A A DNA pool; the minor alleles of 524 were observed within the bovine strain *tir* 255T > A T DNA pool, and the minor alleles of 434 were observed within the human strain *tir* 255T > A T DNA pool, with average frequencies of 18.3%, 23.9%, and 23.9%, respectively. Six hundred and twenty-nine of the 762 polymorphisms reside in open reading frames with 379 predicted nonsynonymous or premature stop codon allele variants and 250 predicted synonymous allele variants ([supplementary table 2, Supplementary Material online](#)).

Identification and Distribution of Polymorphism-Derived Genotypes in STEC O157 Strains Isolated from Humans or Cattle

A total of 175 unique polymorphism-derived genotypes were identified from the 426 STEC O157 strains ([fig. 3,](#)

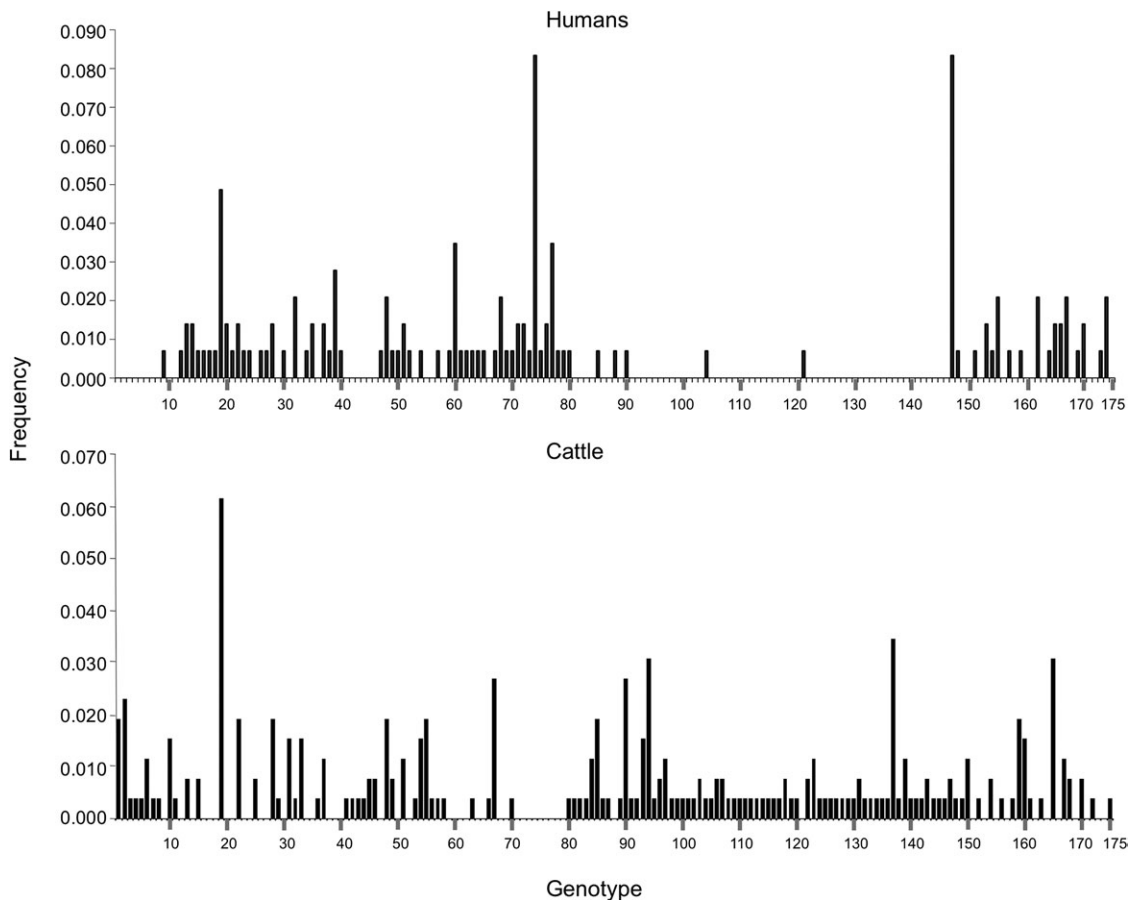


Fig. 3. Frequencies of 175 polymorphism-derived genotypes in STEC O157 strains isolated from humans and cattle. Individual strain genotypes and genotype sequences can be found in [supplementary tables 1 and 3, Supplementary Material](#) online, respectively.

[supplementary tables 3 and 4, Supplementary Material](#) online). Of those, 76 were observed in strains isolated from humans and 124 were observed in strains isolated from cattle. Ninety-eight genotypes were observed exclusively in cattle-isolated strains, 50 in human-isolated strains, and 26 were observed in both human- and cattle-isolated strains. A relatively frequent genotype from human-isolated strains (genotype #74, frequency = 0.083) was not observed in any of the cattle-isolated strains (fig. 3). This genotype is one of nine (71–79) that were observed in 30 SOR⁺ STEC O157:H⁻ strains that were isolated from humans in Germany and one SOR⁻ strain that was also isolated from a human in Germany. STEC O157 strains isolated from cattle in Germany were not represented in this study. However, SOR⁺ STEC O157:H⁻ strains have not been isolated from German cattle to our knowledge (Alpers et al. 2009), and only a few of these strains have been isolated from cattle in Europe and Korea to date (Bielaszewska et al. 2000; Lee and Choi 2006; Orth et al. 2006). Thus, the frequency differences of genotype #74 in cattle- and human-isolated strains are likely due to cattle not being a primary reservoir of SOR⁺ STEC O157:H⁻. In contrast, genotype #147 was observed in 12 STEC O157:H7 strains isolated from humans in Germany (frequency in human-isolated strains = 0.083) and two STEC O157:H7 strains isolated from cattle, one from the United States and one from Scotland (frequency in cattle-isolated

strains = 0.008). The frequency distribution differences of this genotype in cattle- and human-isolated strains are a likely result of geographical stratification of this subtype in Germany combined with an absence of STEC O157 strains isolated from German cattle in our sample collection. All 175 polymorphism-derived genotypes can be resolved with a minimal set of 138 tagging polymorphisms.

Phylogenetic Analysis of Polymorphism-Derived Genotypes and Identification of Eight Major Lineages

Neighbor joining trees were generated from the polymorphism-derived genotypes of all 762 nucleotide polymorphism alleles (fig. 4) and from only the 138 tagging polymorphism alleles (fig. 5). Both trees were rooted with a homologous genotype from STEC O55:H7 and used to track the alleles of the *tir* 255 T > A polymorphism, SOR fermentation, and genotype frequencies in strains isolated from humans. Both trees yielded similar topologies; however, as expected, the bootstrap support was reduced in the phylogenetic tree constructed from the tagging polymorphism alleles as it contained reduced phylogenetic signal (Clawson et al. 2009). The full-length polymorphism-derived genotype tree in figure 4 depicts strongly supported clades that represent several lines of descent within the STEC O157 serogroup. A clade on the tree that branches closely

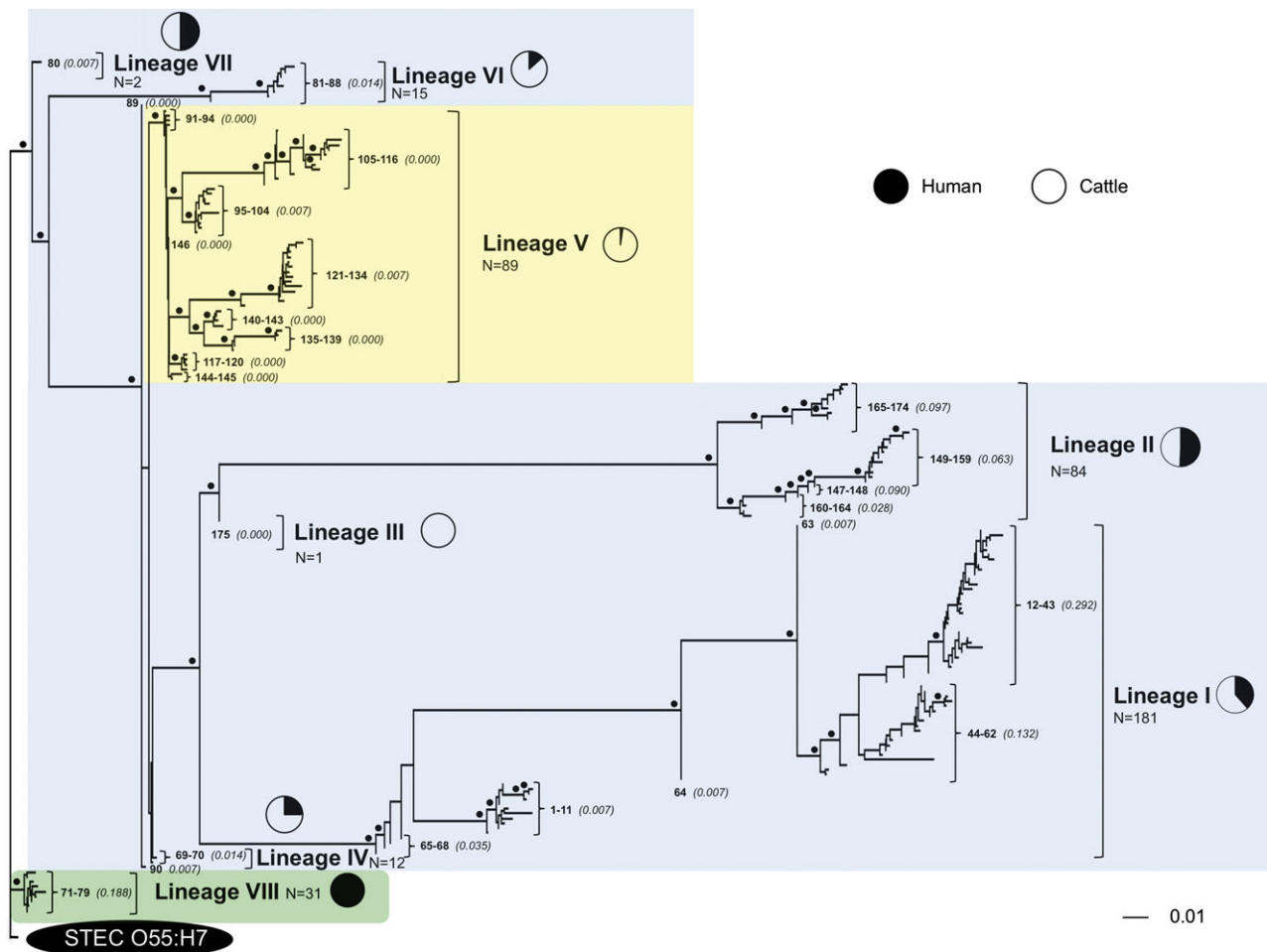


Fig. 4. Neighbor joining tree of full-length STEC O157 polymorphism-derived genotypes. Bullets represent bootstrap values equal to or greater than 80% ($n = 1,000$ bootstraps). Numbers in bold correspond to genotype numbers. Italicized numbers in parentheses represent clade frequencies in the human-isolated strains. Major lineages within the tree are denoted with Roman numerals. Lineages I–IV, VI, and VII are represented by SOR^- , *tir* 255T > A T strains. Lineage V is represented by SOR^- , *tir* T > A A strains. Lineage VIII is represented by SOR^+ , *tir* 255T > A T strains and one mutant SOR^- , *tir* 255T > A T strain. Proportions of human and cattle STEC O157 strains per lineage are represented with black and white pie charts. The numbers of strains per lineage are shown with 'N' counts. Genotype #89 is part of Lineage IV in this tree. The scale bar represents substitutions per site.

to the STEC O55:H7 root is composed of *tir* 255 T > A T, SOR^+ strains. This clade represents polymorphism-derived genotypes from all SOR^+ human-isolated strains in this study ($n = 30$) and one atypical SOR^- fermenting human-isolated strain (Lineage VIII). Genotype #80 represents the early divergence of a SOR^- STEC O157 lineage on the tree and was identified from a human-isolated strain and a cattle-isolated strain, both from the United States (Lineage VII). Genotypes #81-88 also place in a clade that diverged early in the evolution of STEC O157 and were identified in 12 cattle-isolated strains from the United States, 1 cattle-isolated strain from Japan, and 2 human-isolated strains from Germany (Lineage VI). Notably, a strongly supported monophyletic clade of polymorphism-derived genotypes with the *tir* 255T > A A allele appears on the tree after the emergence of SOR^- STEC O157. The clade is represented by 87 cattle-isolated strains that collectively are from the United States, Japan, and Australia and two human-isolated strains, one from the United States and one from Germany (Lineage V).

Thus, this clade is commonly represented in cattle and rarely in humans.

The majority of human-isolated strains placed in one of two major lineages (I and II) that each contain well-defined subclades. Lineage I contains genotypes #1–68, which were identified in 71 of the 154 human-isolated strains genotyped in our sample collection ($n = 7$ from Germany and $n = 64$ from USA). Lineage II, which is represented by more tightly clustered clades on the tree, contains genotypes #147–174, which were identified from 42 human-isolated strains ($n = 17$ from Germany and $n = 25$ from USA). These two lineages are both well represented by cattle-isolated strains and have been previously identified at lower resolution using a smaller set of nucleotide polymorphisms (Clawson et al. 2009).

Two additional lineages (III and IV) are demarcated on the tree. Lineage III shares a distant common ancestor with Lineage II and is represented by a genotype (#175) that was observed from a single ground beef-isolated strain from the

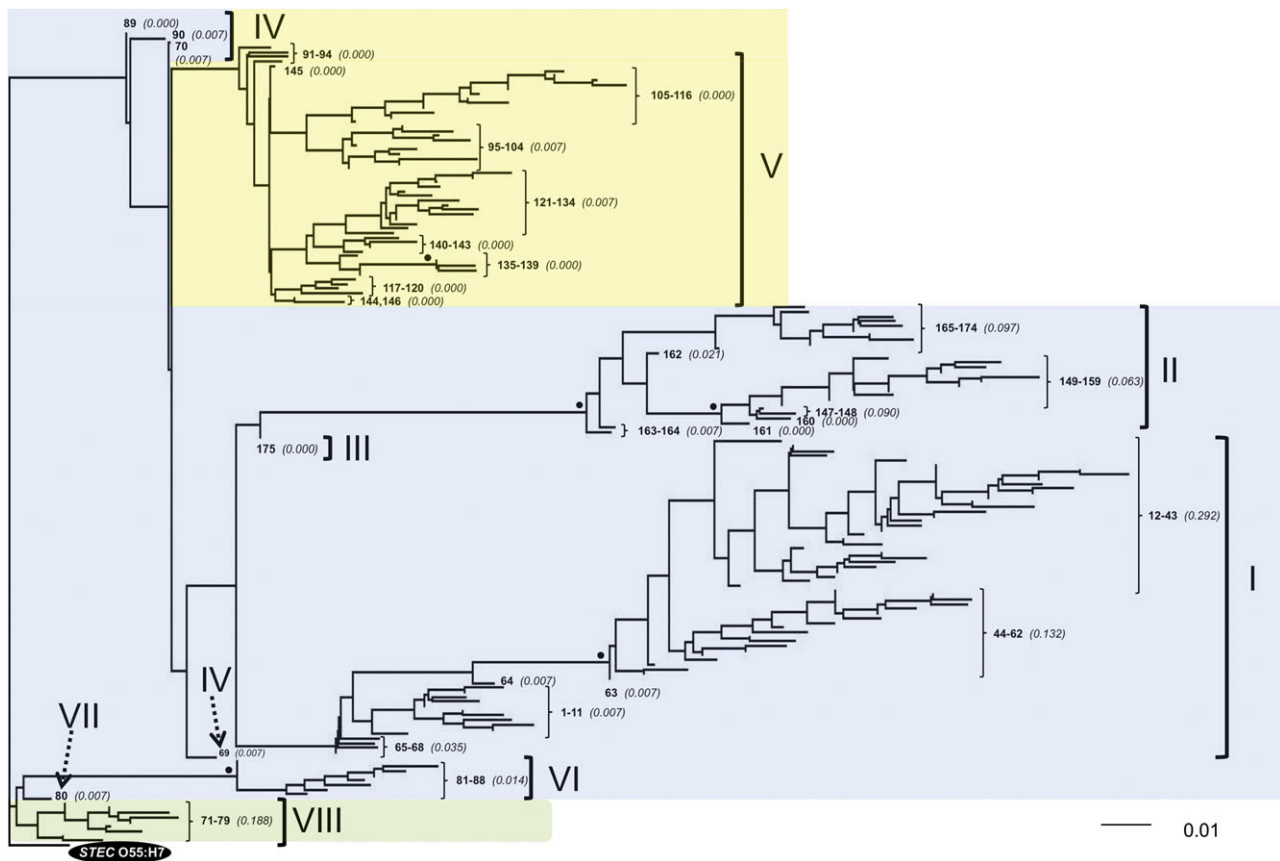


Fig. 5. Neighbor joining tree of STEC O157 polymorphism-derived genotypes based on 138 tagging polymorphisms. Bullets represent bootstrap values equal to or greater than 80% ($n = 1,000$ bootstraps). Numbers in bold correspond to genotype numbers. Italicized numbers in parentheses represent clade frequencies in the human strains. Major lineages determined with full-length polymorphism-derived genotypes are denoted with Roman numerals. The scale bar represents substitutions per site.

United States. Lineage IV is composed of a loose cluster of short branch length genotypes (#69–70, #89–90). These genotypes were observed in three human-isolated strains ($n = 2$ from United States and $n = 1$ from Germany) and nine cattle-isolated strains ($n = 1$ from Scotland, $n = 1$ from Australia, and $n = 7$ from United States).

Comparison of SBI Genotypes with Polymorphism-Derived Genotype Phylogeny

To ultimately compare the phylogeny of STEC O157 based on polymorphism-derived genotypes with the earlier model of evolution depicted in figure 1, we had to connect SBI genotypes with the polymorphism-derived genotypes identified in this study. Accordingly, SBI genotype scores were obtained from 146 human-isolated STEC O157 strains and 230 cattle-isolated strains that were also scored by polymorphism-derived genotypes. Within this collection of strains, two unique SBI genotypes were observed from human-isolated strains and four were observed from cattle-isolated strains (fig. 6). Three SBI genotypes accounted for 89% of the overall genotype frequency in the human-isolated strains (genotypes #1, 3, and 11). One of these (genotype #11) was observed in all 30 SOR⁺ STEC O157H⁻ human-isolated strains and also from one SOR⁻ human-isolated strain. The other two SBI genotypes of high

frequencies in human-isolated strains (genotype #1 = 0.291, #3 = 0.392) were previously reported as high frequency SBI genotypes for the large portion of human-isolated STEC O157 strains in our collection that were scored and analyzed in other studies (Besser et al. 2007; Whitworth et al. 2008), with some geographical variation in their proportional frequencies (Whitworth et al. 2008). Similarly, the high frequencies of four SBI genotypes from cattle-isolated strains in our collection (genotypes #1, 3, 5, and 6; fig. 6) have also been previously reported (Besser et al. 2007; Whitworth et al. 2008).

Placement of SBI Genotypes onto a Phylogenetic Tree of Polymorphism-Derived Genotypes

A rooted phylogenetic tree of polymorphism-derived genotypes was used to connect SBI genotypes with the phylogeny of their corresponding polymorphism-derived genotypes (fig. 7). SBI genotypes are subject to homoplasy due to phage excisions and/or insertions that may be difficult to detect. However, one or two SBI genotypes accounted for at least 50% of the SBI genotype frequency in each of the eight major polymorphism-derived lineages (fig. 7). Thus, the identity of predominant SBI genotypes within the major polymorphism-derived lineages of STEC O157 could be used to compare the polymorphism-derived

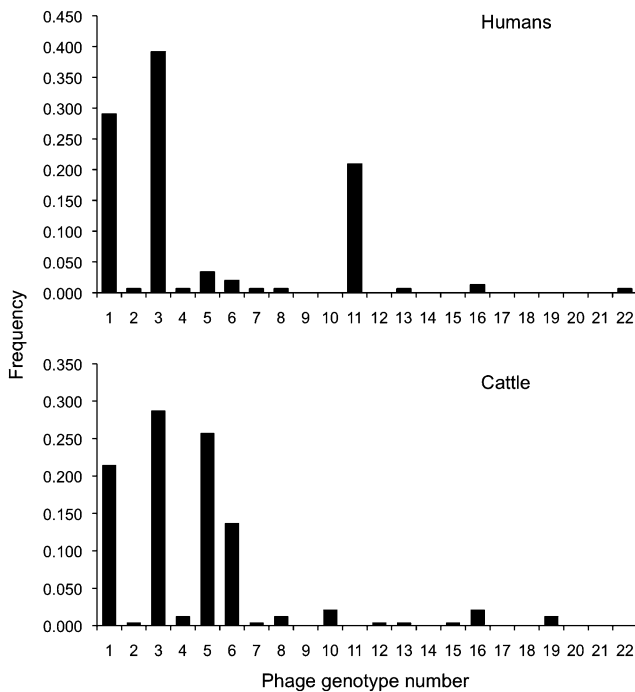


Fig. 6. Frequencies of SBI genotypes in human and cattle STEC O157 strains. SBI genotype numbers represent PCR results (0 for absence and 1 for presence) for the following concatenated genotypes: *stx1*, *stx2*, *yehV*-left and Stx1-encoding bacteriophage junction, *yehV*-right and Stx1-encoding bacteriophage junction, *wrbA*-left and Stx2-encoding bacteriophage junction, and *wrbA*-right and Stx2-encoding bacteriophage junction. Genotype numbers represent the following concatenations: 1 = 011100, 2 = 011111, 3 = 111111, 4 = 010011, 5 = 011000, 6 = 111000, 7 = 111011, 8 = 101100, 10 = 001100, 11 = 010000, 12 = 011011, 13 = 101000, 15 = 110000, 16 = 111100, 19 = 010100, and 22 = 101111.

tree with the current model of STEC O157 evolution represented in [figure 1](#).

Comparison of Eight Major Lineages of Polymorphism-Derived Genotypes with the Current Model of STEC O157 Evolution

Overall, the major polymorphism-derived lineages of STEC O157 support much of the earlier model of STEC O157 evolution, albeit at higher resolution. A comparison of the polymorphism-derived lineages with the earlier model of STEC evolution is shown in [figure 8](#). Lineage VIII contains SOR^+ STEC O157:H⁻ strains that all have the same SBI genotype (genotype #11) and are all $stx2^+$, $stx1^-$, and negative for phage integrations in either *yehV* or *wrbA*. Only one strain of this lineage could be tested for GUD activity and was positive ([supplementary table 1, Supplementary Material online](#)). Consequently, Lineage VIII is similar to subgroup B. Lineage VII contains two strains that are $stx2^+$, SOR^- , and GUD^+ , with one $stx1^+$ strain and one $stx1^-$ strain, and both strains PCR positive for a Stx1-encoding bacteriophage integration spanning the left junction of *yehV* and negative for a Stx2-encoding bacteriophage integration in *wrbA* (SBI genotypes 5 [$stx1^-$] and 6 [$stx1^+$]). Lineage VII is similar to subgroup A5 and in fact contains strain

TW05356, a known subgroup A5 member ([Feng et al. 2007](#)); however, the presence or absence of *stx1* phage integrations in *yehV* was not originally used to define subgroup A5 ([Feng et al. 2007](#)) and one of the strains in Lineage VII is $stx1^-$. Lineage VI is not represented in the early model of STEC O157 evolution and is composed of strains that are all SBI genotype #5. Lineage V, the clade predominantly composed of STEC O157 genotypes from cattle isolates, is also not represented in the early model of STEC O157 evolution. This lineage is composed almost entirely of SBI genotypes #5 and 6. Similarly, SBI genotype #6 is the predominant genotype in Lineages III and IV, both of which are not represented in the early model of STEC O157 evolution.

Lineage II is comprised mainly of strains that are $stx2^+$, $stx1^-$, SOR^- , GUD^- , contain a Stx1-encoding bacteriophage that spans both junctions of *yehV*, and do not contain a Stx2-encoding bacteriophage in *wrbA* (SBI genotype #1). Thus, Lineage II is similar to subgroup C1. However, subgroup C1 is characterized by a truncated Stx1-encoding bacteriophage that occupies the left and right junction of *yehV*. One strain of Lineage II is known to have the exact same truncation of Stx1-encoding bacteriophage as strains that place in subgroup C1 (strain RM 6049, [supplementary table 1, Supplementary Material online](#)), although it is not certain that all other strains that place in Lineage II have this same truncation.

Lineage I contains strains that are similar to subgroups C1, C2, and C3. Constituents of these three subgroups are thought to have sequentially evolved, a model that is fully supported in this study. Strains with SBI genotype #1 are similar to subgroup C1 and place in the basal subclades of Lineage I. The one strain in this study that represents subgroup C2 (strain TW00116, also known as strain 86-24, [supplementary table 1, Supplementary Material online](#)) is $stx1^-$, $stx2^+$, SOR^- , GUD^- , contains a Stx1-encoding bacteriophage that spans both junctions of *yehV* (known to be centrally truncated with the loss of *stx1*), and contains a Stx2-encoding bacteriophage in *wrbA* (SBI genotype #2). Within Lineage I, the polymorphism-derived genotype of this strain (genotype #61) is apical to the subclades connected with SBI genotype #1, and basal to all but two polymorphism-derived genotypes connected with SBI genotype #3, placing SBI genotype #2 as an evolutionary intermediate between the strains with SBI genotype #1 and a majority of strains with SBI genotype #3. Strains with SBI genotype #3 are $stx1^+$, $stx2^+$, SOR^- , GUD^- , contain a Stx1-encoding bacteriophage that spans both junctions of *yehV*, and contain a Stx2-encoding bacteriophage in *wrbA*. The gain of *stx1* in strains with SBI genotype #3 is linked with the gain of an intact Stx1-encoding bacteriophage in *yehV*, and two strains with SBI genotype #3 are known to have an intact Stx1-encoding bacteriophage in *yehV* (strains EDL933 and Sakai, [supplementary table 1, Supplementary Material online](#); [Leopold et al. 2009](#)). These results support the notion that SBI genotype #3, the most frequent SBI genotype observed in this study in both human and cattle-isolated strains, recently emerged during STEC O157 evolution.

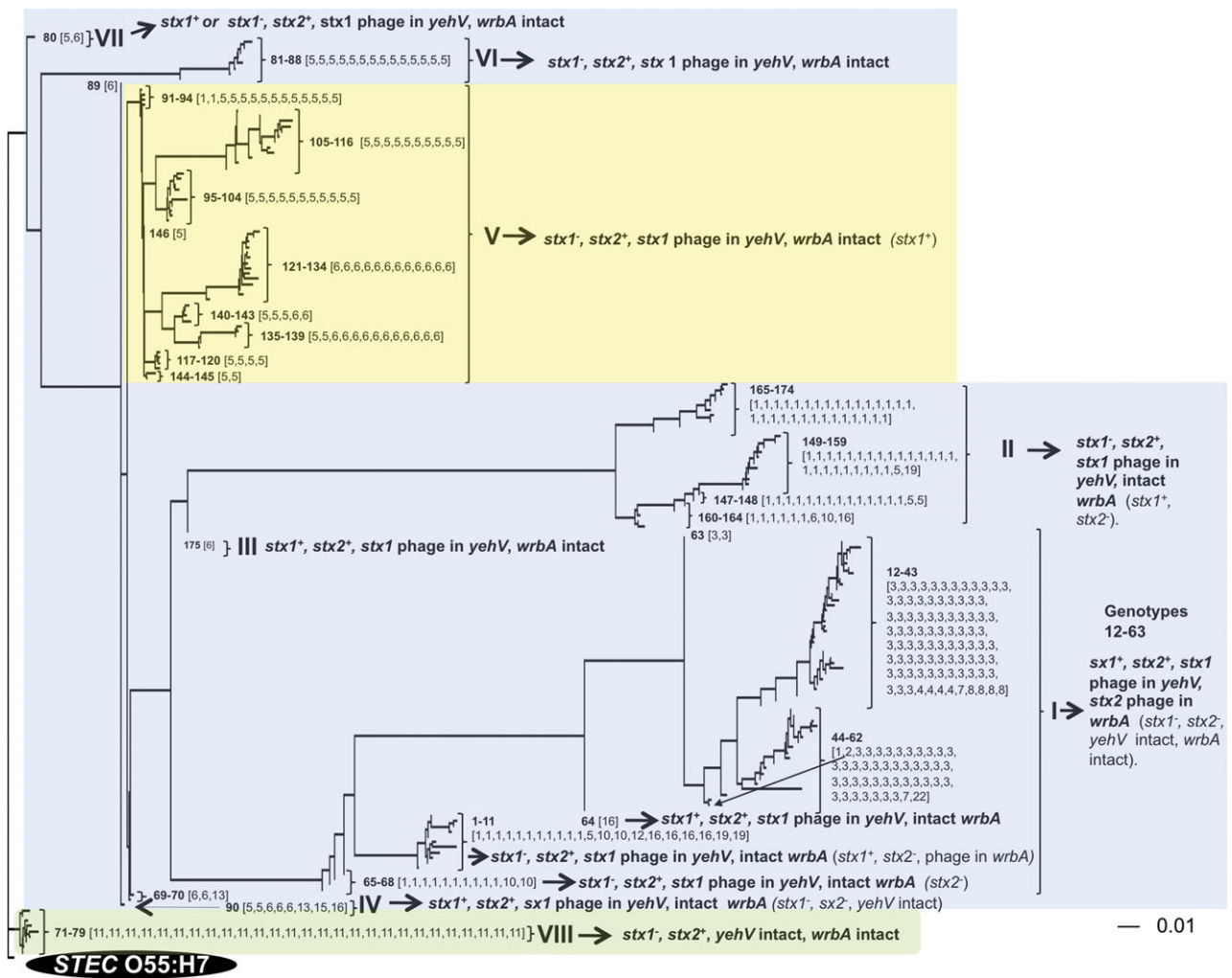


Fig. 7. Neighbor joining tree of full-length STEC O157 polymorphism-derived genotypes with corresponding SBI genotypes. Numbers in bold correspond to polymorphism-derived genotypes. Roman numerals correspond to the major lineages. Numbers in brackets correspond to SBI genotypes and are listed to show their frequencies in polymorphism-derived genotype lineages and/or subclades. SBI genotype traits observed at frequencies of 50% or higher in a lineage or lineage subclade are listed in bold, whereas those observed at less than 50% are in parentheses and are nonbolded. SBI genotypes in Lineage I subclades are shown for clarity. The exact location of a polymorphism-derived genotype corresponding to SBI genotype 2 is denoted on the tree with an arrow as this phage genotype was observed from just one human strain and is the sole representative of the C-2 subgroup. Genotype #89 is part of Lineage IV in this tree. The scale bar represents substitutions per site.

Comparison of PFGE Diversity with That of Polymorphism-Derived Genotypes Defined from 762 Nucleotide Polymorphism Alleles

Genetic diversities based on PFGE and polymorphism-derived genotypes were compared among 96 epidemiologically unrelated STEC O157 strains isolated from ground beef. PFGE profiles were determined through independent *Xba*I and *Bln*I digestions as *Xba*I is typically used as a primary enzyme for determining STEC O157 genetic relatedness and *Bln*I is used as a secondary enzyme to either confirm *Xba*I results or provide additional resolution (Gerner-Smidt et al. 2006). Overall, PFGE diversity assessed with either *Xba*I or *Bln*I surpassed polymorphism-derived genotype diversity. A total of 41 polymorphism-derived genotypes were observed between the strains, whereas *Xba*I and *Bln*I digestions yielded 58 and 50 PFGE patterns, respectively (fig. 9). However, nine *Bln*I PFGE patterns occurred with two or more polymorphism-

derived genotypes, as did 10 *Xba*I PFGE patterns (fig. 9), indicating that the polymorphism-derived genotypes provide added resolution to PFGE banding profiles based on individual digestions in some instances.

PFGE diversity assessed with combined *Xba*I and *Bln*I digestion results surpassed polymorphism-derived genotype diversity to a greater extent than either of the single digestions. Of 68 unique PFGE profiles identified with combined digestion results, 54 were observed once with individual strains, and only 4 of 14 profiles observed from multiple strains occurred with two distinct polymorphism-derived genotypes (fig. 10). Each of the two distinct polymorphism-derived genotypes that occurred with a single PFGE profile are highly related to one another based on their placement as outer taxonomic units in the phylogenetic trees (data not shown). In contrast, 17 polymorphism-derived genotypes occurred with two or

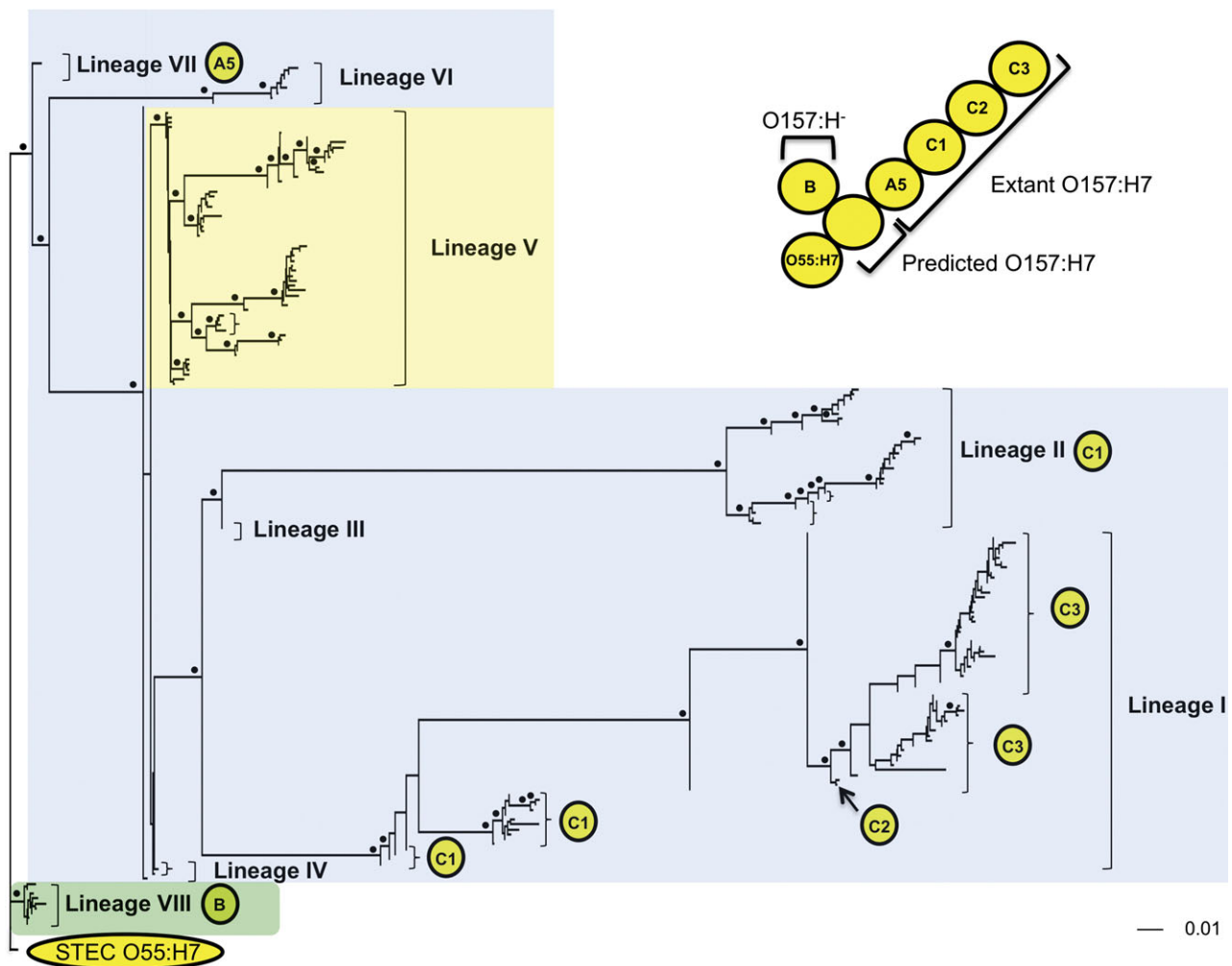


FIG. 8. Comparison of STEC O157 evolution based on polymorphism-derived genotypes with the earlier evolutionary model. The scale bar represents substitutions per site.

more PFGE profiles based on combined digestions. Although the PFGE profiles used in this study do not effectively show genetic relatedness by descent, these results show that PFGE, based on single or combined *Xba*I and *Bln*I digestions, can detect genetic diversity with greater resolution than the polymorphism-derived genotypes defined in this study.

Discussion

As a set, the 762 nucleotide polymorphisms characterized in this study capture a wide spectrum of STEC O157 genetic diversity and were instrumental in identifying eight major lineages of genetic subtypes and determining the order of their descent within the serogroup. However, we caution that the phylogenetic distances (branch lengths) within and between the lineages (figs. 4, 5, 7, and 8) are a reflection of our nucleotide polymorphism selection criteria and the STEC O157 strains used for polymorphism discovery and validation. For example, Lineage VIII, which is composed of SOR^+ STEC O157:H⁻ strain genetic subtypes, has relatively short internal branches in figure 4. This is because

a single SOR^+ STEC O157:H⁻ strain was represented in the nucleotide polymorphism discovery process, and polymorphisms were not selected to delineate this lineage. Consequently, the genetic subtypes of Lineage VIII primarily represent ancestral states for the 762 polymorphisms with minimal phylogenetic signal stemming from individual polymorphism alleles that are specific to the lineage.

An additional example is the common ancestor for Lineage V, which resides on a very short branch despite the selection of 160 polymorphisms for polymorphism-derived genotyping that were specific to the bovine strain *tir* 255T > A A allele DNA pool (fig. 2). The 160 polymorphisms are specific to Lineage V; however, none of them had an allele that occurred at 100% in the bovine strain *tir* 255T > A A allele DNA pool and at 0% frequency in the other two DNA pools. Inclusion of polymorphism alleles of this type would have extended the internal branch leading to the common ancestor of Lineage V, and 181 such putative polymorphisms were identified (Clawson et al. 2009). Thus, the short branch leading to the common ancestor for Lineage V is a result of our selecting 160 polymorphisms that targeted variation between extant genetic subtypes of

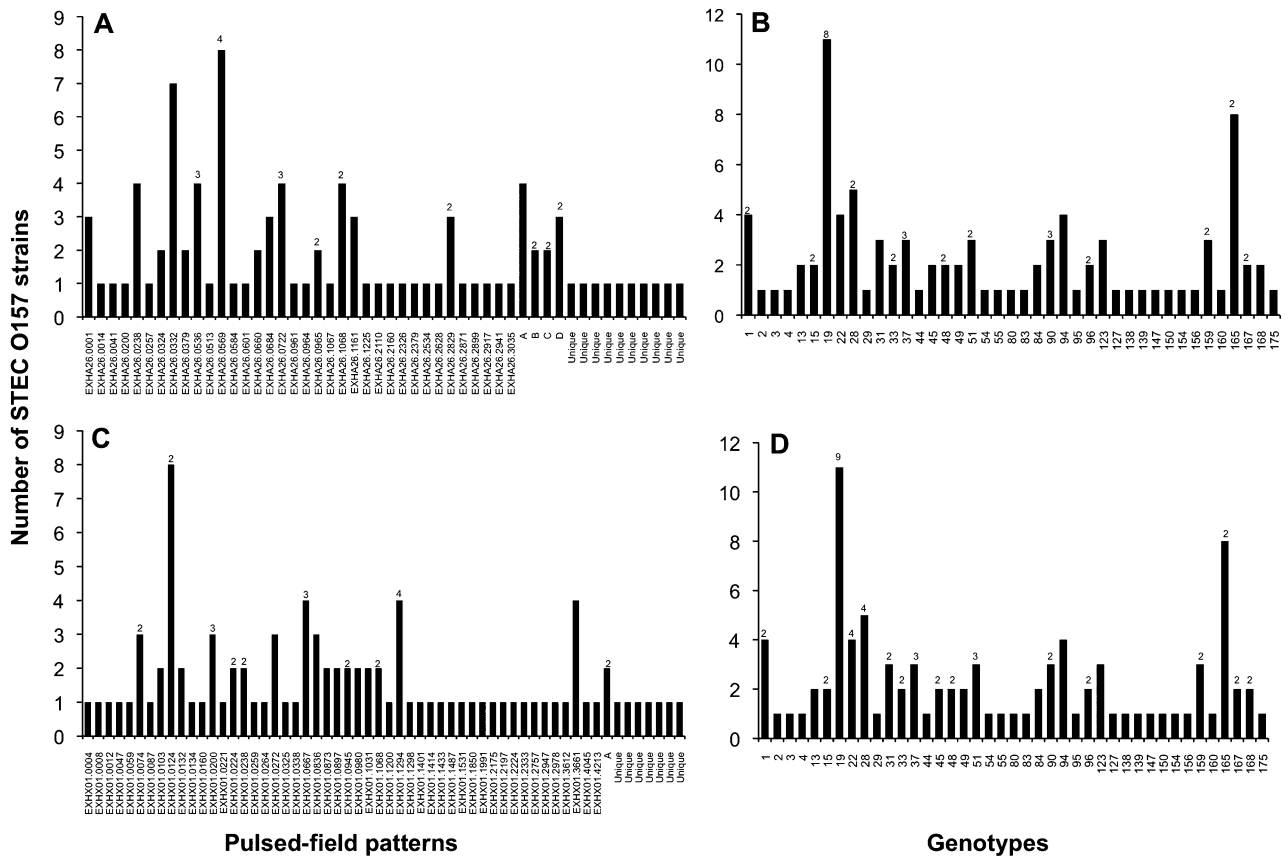


Fig. 9. Comparison of STEC O157 diversity assessed with polymorphism-derived genotypes and independent *XbaI* or *BlnI* PFGE. (A) Number of STEC O157 strains per unique *BlnI* pulsed-field patterns. Numbers above the vertical bars represent the extent of different polymorphism-derived genotypes observed per *BlnI* pulsed-field pattern. (B) Number of STEC O157 strains per polymorphism-derived genotypes. Numbers above the vertical bars represent the extent of different *BlnI* pulsed-field patterns observed per polymorphism-derived genotype. (C) Number of STEC O157 strains per unique *XbaI* pulsed-field pattern. Numbers above the vertical bars represent the extent of different polymorphism-derived genotypes observed per *XbaI* pulsed-field pattern. (D) Number of STEC O157 strains per polymorphism-derived genotypes. Numbers above the vertical bars represent the extent of different *XbaI* pulsed-field patterns observed per genotype. PFGE patterns labeled “A,” “B,” “C,” “D,” or “unique” were not observed in the Pulsenet USA STEC O157 database.

the lineage rather than variation that defined the lineage ancestry. This makes time of divergence calculations for Lineage V and the other lineages problematical. Despite the unequal selection of informative nucleotide polymorphisms across lineages, the phylogenetic signal inherent in this set of nucleotide polymorphisms points to a strong order of descent within the STEC O157 serogroup that is well supported by bootstraps and maintained using a minimal set of 138 tagging polymorphisms.

Little evidence exists that indicates cattle are a reservoir for STEC O55:H7 or STEC O157:H⁻ (the first lineage to diverge within the STEC O157 serogroup [Lineage VII]). *E. coli* strains of the ancestral O55 serogroup and STEC O157:H⁻ have only been found infrequently in cattle (Bielaszewska et al. 2000; Kobayashi et al. 2001; Wani et al. 2003; Lee and Choi 2006). Notably, humans are a primary host for *E. coli* O55:H7, which are a worldwide cause of infantile diarrhea (Whittam et al. 1993; Spears et al. 2006), whereas the primary reservoir for SOR⁺ STEC O157:H⁻ remains unknown (Alpers et al. 2009). Given that cattle-isolated strains exceeded human-isolated strains in this study and Lineages III and VII were represented by 1 and 2 strains overall, re-

spectively (fig. 4, supplementary table 1, Supplementary Material online), our cattle strain genetic subtypes placed into STEC O157 Lineages I-VII and accounted for 50% or greater of the subtype numbers observed in Lineage I, and III–VII, and 45% of Lineage II. Thus, cattle maintain their reservoir status with a preponderance of genetic subtypes that evolved after the divergence of Lineage VIII. Where and how cattle originally acquired STEC O157 is unknown. One possibility is humans; however, it is also possible that cattle once were a reservoir for STEC O55:H7 and Lineage VIII. Furthermore, that STEC O157 evolved from an STEC O55:H7 progenitor in cattle and subsequently diversified, followed by cattle then losing the ability to be a reservoir for STEC O55:H7 and Lineage VIII. Additionally, STEC O157 have been found sporadically in domestic animals, synanthropic rodents, birds, amphibians, fish, insects, mollusks, and feral swine (Jay et al. 2007; Ferens and Hovde 2011). Deer also have been found to harbor STEC O157 at low frequencies (Renter et al. 2001; Rice et al. 2003), including strains that appear to have ancestral characteristics (García-Sánchez et al. 2007). Thus, in addition to humans, cattle may have originally acquired STEC O157 from deer or other wildlife.

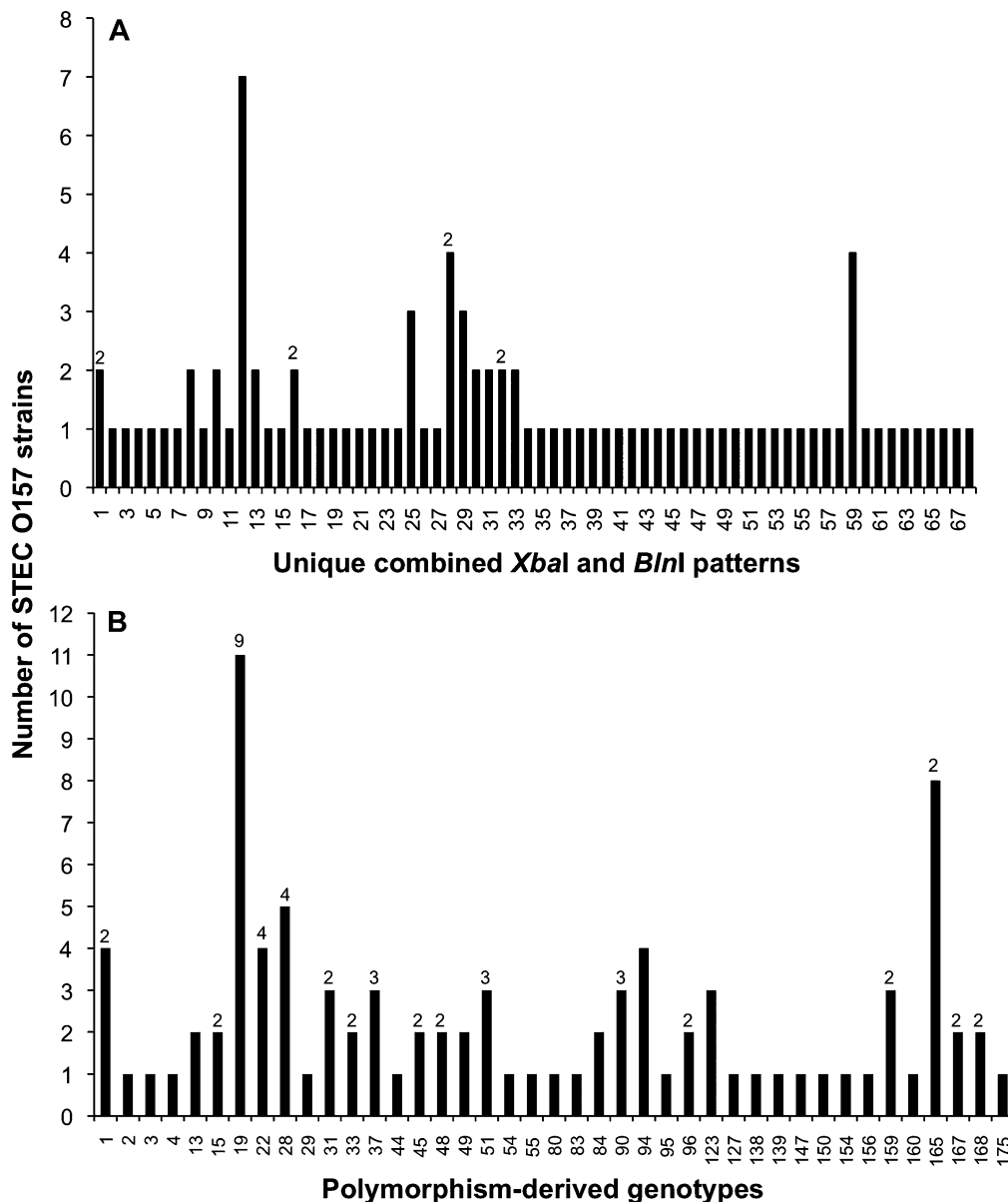


Fig. 10. Comparison of STEC O157 diversity assessed with polymorphism-derived genotypes and combined *XbaI* and *BlnI* PFGEs. (A) Number of STEC O157 strains observed per unique combined *XbaI* and *BlnI* PFGE patterns. Numbers above the bars represent the extent of different polymorphism-derived genotypes observed per combined PFGE pattern. Corresponding PulseNet USA STEC O157 database *XbaI* and *BlnI* PFGE patterns for each of the 68 unique combinations are listed in [supplementary table 1, Supplementary Material](#) online. (B) Number of STEC O157 strains observed per polymorphism-derived genotype. Numbers above the vertical bars represent the extent of different unique PFGE combinations observed per genotype.

As Lineage V genetic subtypes are maintained by cattle and rarely found in clinically ill humans, an ancestor for this lineage may have evolved away from human virulence and disseminated the trait to extant members of the lineage. The identification of genetic mechanisms that may cause the underrepresentation of lineage V with human disease is a matter of intense interest. The *tir* 255T > A A allele is an intrinsic character for Lineage V and, thus, a likely suspect for impacting human virulence. However, the T > A mutation, which encodes a nonsynonymous replacement of aspartate for glutamate within the translocated receptor protein, has not been shown to directly affect human virulence (Bono et al. 2007). The 160 validated polymor-

phisms that were observed exclusively from the bovine strain *tir* 255T > A A allele DNA pool could have an impact on human virulence, especially since three of them resulted in premature stop codons and 80 coded for nonsynonymous mutations. However, these polymorphisms evolved on the backbone of the *tir* 255T > A A allele and were not shared with the last common ancestor of Lineage V. To that end, the genome of a *tir* 255T > A A STEC O157 strain was sequenced and compared with the whole-genome sequences of two other *tir* 255T > A A STEC O157 strains and those of *tir* 255T > A T strains (Eppinger et al. 2011). The study identified 298 polymorphisms that, within a phylogenetic tree, defined a long

internal branch leading to extant genetic subtypes of the *tir* 255T > A lineage. These polymorphisms represent a large number of plausible genetic variants that could have a biological role in reduced human virulence.

Additionally, the Stx complement of Lineage V may very well have a role in reduced human virulence. Stx1 and Stx 2 are important pathogenic determinants of both STEC O157 and non-O157 serogroups as they are the central cause of hemorrhagic colitis and HUS (Law 2000). Stx1 and Stx2 are both AB holotoxins; however, as a cytotoxic agent, Stx2 is 1,000 times more potent to human renal microvascular endothelial cells than Stx1 (Louise and Obrig 1995). At least 12 *stx2* subtypes have been identified across the STEC O157 and non-O157 serogroups that differ in their association with human disease severity (Bertin et al. 2001; Persson et al. 2007; Kawano et al. 2008, 2012). Our phage typing system does not identify phage subtypes. However, all genetic subtypes that place in Lineage V are positive for *stx2*; thus, it is entirely possible that a particular *stx2* subtype is overrepresented in Lineage V, and either by itself or in linkage with polymorphism alleles that are specific to Lineage V, is responsible for reduced human virulence.

The set of 138 tagging polymorphisms developed in this study that resolve all 175 polymorphism-derived genotypes shows great potential in detecting STEC O157 genetic subtypes and determining their evolutionary relatedness. However, PFGE remains an effective tool for detecting variation between closely related strains. PFGE and polymorphism-derived genotyping are complementary methods as PFGE primarily detects insertion/deletion variation within genomic regions specific to STEC O157 (Kudva et al. 2002), whereas the polymorphism set primarily targets a 4.1 Mb backbone conserved among *E. coli* serogroups. Although combined *Xba*I and *Bln*I PFGE patterns revealed greater genetic diversity than the polymorphism-derived genotypes identified in this study, the combined patterns are not sufficient to infer genetic relationships in the absence of epidemiological data (Davis, Hancock, Besser, and Call 2003). Thus, polymorphism-derived genotyping combined with PFGE could be very useful in assessing strain diversity and evolutionary relatedness between epidemiologically unrelated strains.

In summary, the alleles of 762 polymorphisms were used to infer the evolution of STEC O157 strains isolated from cattle and clinically ill humans. Eight major lineages of STEC O157 were identified, of which cattle are known to be a reservoir for seven. One lineage of STEC O157 subtypes that is found in cattle and rarely in humans appears to have evolved away from an association with human disease. All 175 STEC O157 genetic subtypes observed in this study are tagged with a minimal set of 138 polymorphisms. This set can directly assess the evolutionary relatedness of STEC O157 genetic subtypes regardless of their epidemiology.

Supplementary Material

Supplementary tables 1–5 are available at *Molecular Biology and Evolution* online (<http://www.mbe.oxfordjournals.org/>).

Acknowledgments

We thank Gennie Schuller-Chavez, Sandy Fryda-Bradley, Renee Godtel, Linda Flathman, Bob Lee, Kevin Tennill, and Jacky Carnahan for outstanding technical support for this project; Jim Wray, Phil Anderson, and Randy Bradley for computer support; and Joan Rosch for secretarial support. The use of product and company names is necessary to accurately report the methods and results; however, the United States Department of Agriculture (USDA) neither guarantees nor warrants the standard of the products, and the use of names by the USDA implies no approval of the product to the exclusion of others that may also be suitable. The USDA is an equal opportunity provider and employer. This work was supported by the Agricultural Research Service to J.L.B., T.P.L.S., G.P.H., T.G.M., R.E.M., and M.L.C. This research was conducted at the USDA, Agricultural Research Service (ARS), US Meat Animal Research Center (USMARC), State Spur 18D, Clay Center, Nebraska, United States.

References

- Alpers K, Werber D, Frank C, et al. (15 co-authors). 2009. Sorbitol-fermenting enterohaemorrhagic *Escherichia coli* O157:H⁻ causes another outbreak of haemolytic uraemic syndrome in children. *Epidemiol Infect.* 137:389–395.
- Barrett JC, Fry B, Maller J, Daly MJ. 2005. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 21:263–265.
- Bertin Y, Boukhors K, Pradel N, Livrelli V, Martin C. 2001. Stx2 subtyping of Shiga toxin-producing *Escherichia coli* isolated from cattle in France: detection of a new Stx2 subtype and correlation with additional virulence factors. *J Clin Microbiol.* 39:3060–3065.
- Besser TE, Shaikh N, Holt NJ, Tarr PI, Konkel ME, Malik-Kale P, Walsh CW, Whittam TS, Bono JL. 2007. Greater diversity of Shiga toxin-encoding bacteriophage insertion sites among *Escherichia coli* O157:H7 isolates from cattle than in those from humans. *Appl Environ Microbiol.* 73:671–679.
- Bielaszewska M, Prager R, Zhang W, Friedrich AW, Mellmann A, Tschäpe H, Karch H. 2006. Chromosomal dynamism in progeny of outbreak-related sorbitol-fermenting enterohaemorrhagic *Escherichia coli* O157:NM. *Appl Environ Microbiol.* 72:1900–1909.
- Bielaszewska M, Schmidt H, Liesegang A, Prager R, Rabsch W, Tschäpe H, Cizek A, Janda J, Bláhová K, Karch H. 2000. Cattle can be a reservoir of sorbitol-fermenting shiga toxin-producing *Escherichia coli* O157:H(-) strains and a source of human diseases. *J Clin Microbiol.* 38:3470–3473.
- Bono JL, Keen JE, Clawson ML, Durso LM, Heaton MP, Laegreid WW. 2007. Association of *Escherichia coli* O157:H7 *tir* polymorphisms with human infection. *BMC Infect Dis.* 7:98.
- Bono JL, Keen JE, Miller LC, Fox JM, Chitko-McKown CG, Heaton MP, Laegreid WW. 2004. Evaluation of a real-time PCR kit for detecting *Escherichia coli* O157 in bovine fecal samples. *Appl Environ Microbiol.* 70:1855–1857.
- Borczyk AA, Karmali MA, Lior H, Duncan LM. 1987. Bovine reservoir for verotoxin-producing *Escherichia coli* O157:H7. *Lancet* 1:98.
- Brown PK, Dozois CM, Nickerson CA, Zuppardo A, Terlonge J, Curtis RI III. 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.
- Clawson ML, Keen JE, Smith TPL, Durso LM, McDanel TG, Mandrell RE, Davis MA, Bono JL. 2009. Phylogenetic classification of *Escherichia coli* O157:H7 strains of human and bovine

- origin using a novel set of nucleotide polymorphisms. *Genome Biol.* 10:R56.
- Cooley M, Carychao D, Crawford-Miksza L, Jay MT, Myers C, Rose C, Keys C, Farrar J, Mandrell RE. 2007. Incidence and tracking of *Escherichia coli* O157:H7 in a major produce production region in California. *PLoS One* 2(11):e1159.
- Davis MA, Hancock DD, Besser TE, Call DR. 2003. Evaluation of pulsed-field gel electrophoresis as a tool for determining the degree of genetic relatedness between strains of *Escherichia coli* O157:H7. *J Clin Microbiol.* 41:1843–1849.
- Davis MA, Hancock DD, Besser TE, Rice DH, Hovde CJ, Digiacomo R, Samadpour M, Call DR. 2003. Correlation between geographic distance and genetic similarity in an international collection of bovine faecal *Escherichia coli* O157:H7 isolates. *Epidemiol Infect.* 131:923–930.
- Elder RO, Keen JE, Siragusa GR, Barkocy-Gallagher GA, Koochmaria M, Laegreid WW. 2000. Correlation of enterohemorrhagic *Escherichia coli* O157 prevalence in feces, hides, and carcasses of beef cattle during processing. *Proc Natl Acad Sci U S A.* 97:2999–3003.
- Eppinger M, Mammel MK, LeClerc JE, Ravel J, Cebula TA. 2011. Genome signatures of *Escherichia coli* O157:H7 isolates from the bovine host reservoir. *Appl Environ Microbiol.* 77:2916–2925.
- Feng PCH, Monday SR, Lacher DW, et al. (11 co-authors). 2007. Genetic diversity among clonal lineages within *Escherichia coli* O157:H7 stepwise evolutionary model. *Emerg Infect Dis.* 13:1701–1706.
- Ferens WA, Hovde CJ. 2011. *Escherichia coli* O157:H7: animal reservoir and sources of human infection. *Foodborne Pathog Dis.* 8:465–487.
- Gannon VPJ, D'Souza S, Graham T, King RK, Rahn K, Read S. 1997. Use of the flagellar H7 gene as a target in multiplex PCR assays and improved specificity in identification of enterohemorrhagic *Escherichia coli* strains. *J Clin Microbiol.* 35:656–662.
- García-Sánchez A, Sánchez S, Rubio R, Pereira G, Alonso JM, Hermoso de Mendoza J, Rey J. 2007. Presence of Shiga toxin-producing *E. coli* O157:H7 in a survey of wild artiodactyls. *Vet Microbiol.* 121:373–377.
- Gerner-Smidt P, Hise K, Kincaid J, Hunter S, Rolando S, Hyytiä-Trees E, Ribot EM, Swaminathan B. PulseNet Taskforce. 2006. PulseNet USA: a five-year update. *Foodborne Pathog Dis.* 3:9–19.
- Grandori R, Khalifah P, Boice JA, Fairman R, Giovanielli K, Carey J. 1998. Biochemical characterization of WrbA, founding member of a new family of multimeric flavodoxin-like proteins. *J Biol Chem.* 273:20960–20966.
- Griffin PM, Tauxe RV. 1991. The epidemiology of infections caused by *Escherichia coli* O157:H7, other enterohemorrhagic *E. coli*, and the associated hemolytic uremic syndrome. *Epidemiol Rev.* 13:60–98.
- Gunzer F, Böhm H, Rüssmann H, Bitzan M, Aleksic S, Karch H. 1992. Molecular detection of sorbitol-fermenting *Escherichia coli* O157 in patients with hemolytic-uremic syndrome. *J Clin Microbiol.* 30:1807–1810.
- Hayashi T, Makino K, Ohnishi M, et al. (22 co-authors). 2001. Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genomic comparison with a laboratory strain K-12. *DNA Res.* 8:11–22.
- He Y, Keen JE, Westerman RB, Littledike ET, Kwang J. 1996. Monoclonal antibodies for detection of the H7 antigen of *Escherichia coli*. *Appl Environ Microbiol.* 62:3325–3332.
- Huson DH, Richter DC, Rausch C, DeZulian T, Franz M, Rupp R. 2007. Dendroscope: an interactive viewer for large phylogenetic trees. *BMC Bioinformatics* 8:460.
- Iguchi A, Osawa R, Kawano J, Shimizu A, Terajima J, Watanabe H. 2002. Effects of repeated subculturing and prolonged storage at room temperature of enterohemorrhagic *Escherichia coli* O157:H7 on pulsed-field gel electrophoresis profiles. *J Clin Microbiol.* 40:3079–3081.
- Jackson SA, Mammel MK, Patel IR, Mays T, Albert TJ, LeClerc JE, Cebula TA. 2007. Interrogating genomic diversity of *E. coli* O157:H7 using DNA tiling arrays. *Forensic Sci Int.* 168:183–199.
- Jay MT, Cooley M, Carychao D, et al. (13 co-authors). 2007. *Escherichia coli* O157:H7 in feral swine near spinach fields and cattle, central California coast. *Emerg Infect Dis.* 13:1908–1911.
- Kawano K, Okada M, Haga T, Maeda K, Goto Y. 2008. Relationship between pathogenicity for humans and stx genotype in Shiga toxin-producing *Escherichia coli* serotype O157. *Eur J Clin Microbiol Infect Dis.* 27:227–232.
- Kawano K, Ono H, Iwashita O, Kurogi M, Haga T, Maeda K, Goto Y. 2012. stx genotype and molecular epidemiological analyses of Shiga toxin-producing *Escherichia coli* O157:H7/H- in human and cattle isolates. *Eur J Clin Microbiol Infect Dis.* 31:119–127.
- Keen JE, Wittum TE, Dunn JR, Bono JL, Durso LM. 2006. Shiga-toxigenic *Escherichia coli* O157 in agricultural fair livestock, United States. *Emerg Infect Dis.* 12:780–786.
- Kim J, Nietfeldt J, Benson AK. 1999. Octamer-based genome scanning distinguishes a unique subpopulation of *Escherichia coli* O157:H7 strains in cattle. *Proc Natl Acad Sci U S A.* 96:13288–13293.
- Kobayashi H, Shimada J, Nakazawa M, Morozumi T, Pohjanvirta T, Pelkonen S, Yamamoto K. 2001. Prevalence and characteristics of shiga toxin-producing *Escherichia coli* from healthy cattle in Japan. *Appl Environ Microbiol.* 67:484–489.
- Kudva IT, Evans PS, Perna NT, Barrett TJ, Ausubel FM, Blattner FR, Calderwood SB. 2002. Strains of *Escherichia coli* O157:H7 differ primarily by insertions or deletions, not single-nucleotide polymorphisms. *J Bacteriol.* 184:1873–1879.
- Law D. 2000. Virulence factors of *Escherichia coli* O157 and other shiga toxin-producing *E. coli*. *J Appl Microbiol.* 88:729–745.
- Lee JH, Choi S-J. 2006. Isolation and characteristics of sorbitol-fermenting *Escherichia coli* O157 strains from cattle. *Microbes Infect.* 8:2021–2026.
- Leopold SR, Magrini V, Holt NJ, et al. (15 co-authors). 2009. A precise reconstruction of the emergence and constrained radiations of *Escherichia coli* O157 portrayed by backbone concatenomic analysis. *Proc Natl Acad Sci U S A.* 106:8713–8718.
- Lim JY, Yoon JW, Hovde CJ. 2010. A brief overview of *Escherichia coli* O157:H7 and its plasmid O157. *J Microbiol Biotechnol.* 20:5–14.
- Louise CB, Obrig TG. 1995. Specific interaction of *Escherichia coli* O157:H7-derived shiga-like toxin II with human renal endothelial cells. *J Infect Dis.* 172:1397–1401.
- Manning SD, Motiwala AS, Springman AC, et al. (14 co-authors). 2008. Variation in virulence among clades of *Escherichia coli* O157:H7 associated with disease outbreaks. *Proc Natl Acad Sci U S A.* 105:4868–4873.
- Mead PS, Griffin PM. 1998. *Escherichia coli* O157:H7. *Lancet* 352:1207–1212.
- Monday SR, Whittam TS, Feng PCH. 2001. Genetic and evolutionary analysis of mutations in the *gusA* gene that cause the absence of β -glucuronidase activity in *Escherichia coli* O157:H7. *J Infect Dis.* 184:918–921.
- Olsvik O, Rimstad E, Hornes E, Strockbine N, Wasteson Y, Lund A, Wachsmuth K. 1991. A nested PCR followed by magnetic separation of amplified fragments for detection of *Escherichia coli* shiga-like toxin genes. *Mol Cell Probes.* 5:429–435.
- Orth D, Grif K, Dierich MP, Würzner R. 2006. Sorbitol-fermenting shiga toxin-producing *Escherichia coli* O157: indications for an animal reservoir. *Epidemiol Infect.* 134:719–723.
- Paton AW, Paton JC. 1998. Detection and characterization of Shiga toxin-producing *Escherichia coli* by using multiplex PCR assays for *stx*₁, *stx*₂, *eaeA*, enterohemorrhagic *E. coli hlyA*, *rfb*_{O111}, and *rfb*_{O157}. *J Clin Microbiol.* 36:598–602.

- Perna NT, Plunkett G III, Burland V, et al. (28 co-authors). 2001. Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7. *Nature* 409:529–533.
- Persson S, Olsen KEP, Ethelberg S, Scheutz F. 2007. Subtyping method for *Escherichia coli* Shiga toxin (verocytotoxin) 2 variants and correlations to clinical manifestations. *J Clin Microbiol.* 45:2020–2024.
- Pollock KGJ, Locking ME, Beattie TJ, Maxwell H, Ramage I, Hughes D, Cowieson J, Allison L, Hanson M, Cowden JM. 2010. Sorbitol-fermenting *Escherichia coli* O157, Scotland. *Emerg Infect Dis.* 16:881–882.
- Reid SD, Betting DJ, Whittam TS. 1999. Molecular detection and identification of intimin alleles in pathogenic *Escherichia coli* by multiplex PCR. *J Clin Microbiol.* 37:2719–2722.
- Renter DG, Sargeant JM, Hygnstorm SE, Hoffman JD, Gillespie JR. 2001. *Escherichia coli* O157:H7 in free-ranging deer in Nebraska. *J Wildl Dis.* 37:755–760.
- Ribot EM, Fair MA, Gautom R, Cameron DN, Hunter SB, Swaminathan B, Barrett TJ. 2006. Standardization of pulsed-field gel electrophoresis protocols for the subtyping of *Escherichia coli* O157:H7, *Salmonella*, and *Shigella* for PulseNet. *Foodborne Pathog Dis.* 3:59–67.
- Rice DH, Hancock DD, Besser TE. 2003. Faecal culture of wild animals for *Escherichia coli* O157:H7. *Vet Rec.* 152:82–83.
- Roldgaard BB, Scheutz F, Boel J, Aabo S, Schultz AC, Cheasty T, Nielsen EM, Olsen KE, Christensen BB. 2004. VTEC O157 subtypes associated with the most severe clinical symptoms in humans constitute a minor part of VTEC O157 isolates from Danish cattle. *Int J Med Microbiol.* 294:255–259.
- Schmidt HA, Strimmer K, Vingron M, von Haeseler A. 2002. TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics* 18:502–504.
- Shaikh N, Holt NJ, Johnson JR, Tarr PI. 2007. *Fim* operon variation in the emergence of Enterohemorrhagic *Escherichia coli*: an evolutionary and functional analysis. *FEMS Microbiol Lett.* 273:58–63.
- Shaikh N, Tarr PI. 2003. *Escherichia coli* O157:H7 Shiga toxin-encoding bacteriophages: integrations, excisions, truncations, and evolutionary implications. *J Bacteriol.* 185:3596–3605.
- Spears KJ, Roe AJ, Gally DL. 2006. A comparison of enteropathogenic and enterohaemorrhagic *Escherichia coli* pathogenesis. *FEMS Microbiol Lett.* 255:187–202.
- Tarr PI, Schoening LM, Yea Y-L, Ward TR, Jelacic S, Whittam TS. 2000. Acquisition of the *rfb-gnd* cluster in evolution of *Escherichia coli* O55 and O157. *J Bacteriol.* 182:6183–6191.
- Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* 25:4876–4882.
- Wani SA, Bhat MA, Samanta I, Nishikawa Y, Buchh AS. 2003. Isolation and characterization of shiga toxin-producing *Escherichia coli* (STEC) and enteropathogenic *Escherichia coli* (EPEC) from calves and lambs with diarrhoea in India. *Lett Appl Microbiol.* 37:121–126.
- Wells JG, Shipman LD, Greene KD, et al. (13 co-authors). 1991. Isolation of *Escherichia coli* serotype O157:H7 and other shiga-like-toxin-producing *E. coli* from dairy cattle. *J Clin Microbiol.* 29:985–989.
- Westerman RB, He Y, Keen JE, Littledike ET, Kwang J. 1997. Production and characterization of monoclonal antibodies specific for the lipopolysaccharide of *Escherichia coli* O157. *J Clin Microbiol.* 35:679–684.
- Whittam TS. 1998. Evolution of *Escherichia coli* O157. H7 and other shiga toxin-producing *E. coli* strains. In: Kaper JB, O'Brien AD, editors. *Escherichia coli* O157:H7 and other shiga toxin-producing *E. coli* strains. Washington (DC): ASM Press. p. 195–209.
- Whittam TS, Wolfe ML, Wachsmuth IK, Ørskov F, Ørskov I, Wilson RA. 1993. Clonal relationships among *Escherichia coli* strains that cause hemorrhagic colitis and infantile diarrhea. *Infect Immun.* 61:1619–1629.
- Whitworth J, Zhang Y, Bono J, Pleydell E, French N, Besser T. 2010. Diverse genetic markers concordantly identify bovine origin *Escherichia coli* O157 genotypes underrepresented in human disease. *Appl Environ Microbiol.* 76:361–365.
- Whitworth JH, Fegan N, Keller J, Gobius KS, Bono JL, Call DR, Hancock DD, Besser TE. 2008. International comparison of clinical, bovine, and environmental *Escherichia coli* O157 isolates on the basis of shiga toxin-encoding bacteriophage insertion site genotypes. *Appl Environ Microbiol.* 74:7447–7450.
- Zhang W, Qi W, Albert TJ, et al. 2006. Probing genomic diversity and evolution of *Escherichia coli* O157 by single nucleotide polymorphisms. *Genome Res.* 16:757–767.