Genomic Diversity and Virulence Profiles of Historical *Escherichia coli* O157 Strains Isolated from Clinical and Environmental Sources

Lydia V. Rump,^a Narjol Gonzalez-Escalona,^b Wenting Ju,^a Fei Wang,^a Guojie Cao,^a Sean Meng,^a Jianghong Meng^{a,c}

Department of Nutrition & Food Science, University of Maryland, College Park, Maryland, USA^a; Division of Microbiology, Center for Food Safety and Applied Nutrition, Food and Drug Administration, College Park, Maryland, USA^b; Joint Institute for Food Safety and Applied Nutrition, University of Maryland, College Park, Maryland, USA^c

Escherichia coli O157:H7 is, to date, the major *E. coli* serotype causing food-borne human disease worldwide. Strains of O157 with other H antigens also have been recovered. We analyzed a collection of historic O157 strains ($n = 400$) isolated in the late 1980s to early 1990s in the United States. Strains were predominantly serotype O157:H7 (55%), and various O157:non-H7 (41%) serotypes were not previously reported regarding their pathogenic potential. Although lacking Shiga toxin (*stx*) and *eae* genes, serotypes O157:H1, O157:H2, O157:H11, O157:H42, and O157:H43 carried several virulence factors (*iha*, *terD*, and *hlyA*) also found in virulent serotype *E. coli* O157:H7. Pulsed-field gel electrophoresis (PFGE) showed the O157 serogroup was diverse, with strains with the same H type clustering together closely. Among non-H7 isolates, serotype O157:H43 was highly prevalent (65%) and carried important enterohemorrhagic *E. coli* (EHEC) virulence markers (*iha*, *terD*, *hlyA*, and *espP*). Isolates from two particular H types, H2 and H11, among the most commonly found non-O157 EHEC serotypes (O26:H11, O111:H11, O103:H2/H11, and O45:H2), unexpectedly clustered more closely with O157:H7 than other H types and carried several virulence genes. This suggests an early divergence of the O157 serogroup to clades with different pathogenic potentials. The appearance of important EHEC virulence markers in closely related H types suggests their virulence potential and suggests further monitoring of those serotypes not implicated in severe illness thus far.

Enterohemorrhagic *Escherichia coli* (EHEC) strains producing Shiga toxin(s) (Stx) are among the most important food-borne bacterial pathogens, being responsible for outbreaks and severe cases of hemorrhagic colitis (HC) and hemolytic-uremic syndrome (HUS) around the world (1, 2). Over 470 Shiga toxin-producing *E. coli* (STEC) serotypes have been responsible for sporadic and epidemic human infections, although *E. coli* O157:H7 remains the most important and deadly STEC serotype worldwide (3). The main trait of EHEC isolates is their ability to produce and release Stxs, which are considered the major virulence factors of these pathogens. Typical virulent O157:H7 strains carry additional virulence factors besides Stx1 and Stx2, which also contribute to their high pathogenicity, including, for example, the γ -intimin allele (γ -*eae*) on the locus of the enterocyte effacement (LEE) pathogenicity island, which codes for proteins necessary for attaching and effacing lesions of epithelial cells. The EHEC virulence plasmid encodes enterohemolysin (*hlyA*) and other virulence factors (2, 4).

The acquisition of virulence factors may promote the emergence of highly virulent *E. coli* strains from progenitor strains only rarely associated with disease. For example, in 2011, over 4,000 people were sickened by an unusual enteroaggregative *E. coli* (EAEC) O104:H4 strain, which had acquired an *stx*₂ phage and caused a large outbreak of bloody diarrhea with an unusually high number of HUS cases (800) and deaths (53). The mechanisms underlying the evolution and emergence of such new bacterial pathogens are not well understood, and these atypical STEC strains now are attracting more attention (5).

The variety of virulence determinants present may differ between not only serogroups but also individual serotypes due to horizontal gene transfers of mobile elements, such as phages, pathogenicity islands (PAIs), and plasmids contributing to a strain's pathogenic potential (6). Since *stx* genes are located on phages, their gain and loss

frequently alters virulence, leading to highly virulent serotypes, as has been shown for *E. coli* O104:H4 (5).

Phenotypically, *E. coli* O157:H7, the most frequently isolated STEC, is divided into sorbitol-fermenting (SF) nonmotile O157:H strains and non-sorbitol-fermenting (NSF) O157:H7 strains (7). Unlike other *E. coli* serotypes, most typical O157:H7 strains are unable to ferment sorbitol, which serves as a great advantage in their identification and differentiation (2, 8). However, virulent serotypes of O157 are able to ferment sorbitol and can lead to additional challenges for detection, since the NSF trait is used in culture-based identification methods. This underlines the importance of the presence of virulent SF O157:H7 isolates that are more difficult to quickly detect in routine analysis. Additional genetic characteristics further distinguish strains into clonal complexes (CC), allowing the establishment of a proposed evolutionary model for O157:H7 (7, 9). Serotype O157:H7 is postulated to have derived from a common ancestor that gave rise to both O55:H7 strains (belonging to CC A1 and A2) and, further, to a hypothetical O157 intermediary (CC A3) via serotypic changes. SF O157

Received 21 August 2014 Accepted 30 October 2014

Accepted manuscript posted online 7 November 2014

Citation Rump LV, Gonzalez-Escalona N, Ju W, Wang F, Cao G, Meng S, Meng J. 2015. Genomic diversity and virulence profiles of historical *Escherichia coli* O157 strains isolated from clinical and environmental sources. *Appl Environ Microbiol* 81:569–577. doi:10.1128/AEM.02616-14.

Editor: D. W. Schaffner

Address correspondence to Jianghong Meng, jmeng@umd.edu.

Supplemental material for this article may be found at <http://dx.doi.org/10.1128/AEM.02616-14>.

Copyright © 2015, American Society for Microbiology. All Rights Reserved. doi:10.1128/AEM.02616-14

isolates diverge from this group, forming a separate highly clonal group, CC A4 (10), while O157:H7 strains evolved into the hypothetical NSF CC A5, which in turn diverged into two separate lineages: A7, carrying pO157_2, and A6, carrying pO157 and showing the loss of β -glucuronidase activity (9). Presently, A6 strains account for most food-borne illness caused by EHEC worldwide (11).

The O157 serogroup also includes non-H7 strains, most of which have not been implicated in illnesses, which suggests these lack known virulence factors (12). However, enteropathogenic *E. coli* (EPEC) strains of serotype O157:H45 carrying the *eae* gene have been identified; therefore, other H types might carry virulence marker genes and have the potential to be pathogenic (13).

In this study, we examined a large collection of historic O157 isolates (400), collected in the late 1980s and early 1990s, for the prevalence of known O157:H7 virulence genes (*stx*₁, *stx*₂, *eae*, *iha*, *uidA*, and *tir*) located on chromosomes of phages, pathogenicity islands, or O islands, as well as *terD*, which encodes tellurite resistance in OI-43 and OI-48 in typical O157:H7 strains, and the large EHEC plasmid (*hlyA*, *espP*, *katP*, and *toxB*). These isolates are of special interest, since they were collected around the same time as several large outbreaks of O157:H7 that occurred on the west coast in the United States in the early 1990s (14, 15) and Sakai, Japan, in 1996 (16). Analysis of isolates from the same time period will allow us to better understand both diversity within the O157 serogroup and the evolution of known pathogens (such as O157:H7 strains). This might uncover factors that could lead to new, highly pathogenic *E. coli* serotypes, particularly during the time when O157:H7 outbreaks became more frequent (17). The diversity of these isolates was determined using pulsed-field gel electrophoresis (PFGE) and then compared to known *E. coli* strains described in the stepwise evolutionary model for *E. coli* O157:H7 (7).

MATERIALS AND METHODS

Bacterial strains and growth conditions. A collection of historic *E. coli* O157 isolates ($n = 400$) from different sources (foods, animals/animal feces, and patients) and geographical locations in the United States (1988 to 1993) (see Table S1 in the supplemental material) were plated on sorbitol MacConkey agar at 37°C for 18 h to determine their ability to ferment sorbitol. Overnight cultures grown on tryptic soy agar (TSA) at 37°C were boiled for 10 min at 95°C for genotyping and prevalence of virulence genes. The collection has been provided by the University of Maryland, Department of Nutrition and Food Science, College Park, Maryland, USA.

Genotyping. The presence of both somatic O157 and flagellar H7 antigens was determined using multiplex PCR targeting *per* (O157) and *fliC* (H7) genes, respectively (see Table S1 in the supplemental material). A 25- μ l PCR mix contained 1 U of GoTaq DNA polymerase (Promega, Madison, WI), 1 $\times$ polymerase buffer, 3.5 mM MgCl₂, 200 μ M each deoxynucleoside triphosphate (dNTP), 200 nM each primer, and ~ 1 ng of template DNA. PCR conditions were 3 min at 90°C to activate the Taq polymerase, followed by 30 cycles of denaturation at 95°C for 15 s, primer annealing at 58°C for 1 min, and extension at 72°C for 1.5 min, with a final 1-min extension at 72°C. Products were examined on a 1% agarose gel in 1 $\times$ Tris-borate-EDTA buffer. The H type of non-H7 strains was determined by amplifying and sequencing the flagellar structural gene, *fliC*, as previously described (*fliC*_{1F}, 5'-ATGGCACAAGTCATTAATACCAA-3'; *fliC*_{1497R}, 5'-TTAACCTGCAGCAGAGACA-3') (18). BLAST analysis for identifying and matching *fliC* sequences of various H types in GenBank was used to determine the specific H type (18).

Prevalence of virulence genes. Table 1 lists the primers used to test strains for the presence of EHEC virulence factors by PCR. Chromo-

somally encoded virulence markers *stx*₁, *stx*₂, and *eae* and plasmid-encoded *hlyA* were determined using multiplex PCR using PCR reagent using the concentrations and conditions described for genotyping. The remaining virulence markers were detected in individual PCRs with an annealing temperature of 58°C, except for *tir*, *terD*, and *iha* (all at 55°C) and *uidA*, *espP*, *katP*, and *toxB* (all at 60°C).

PFGE. Strains of *E. coli* O157:non-H7 and selected O157:H7 strains carrying virulence genes were characterized by pulsed-field gel electrophoresis (PFGE) using the rapid standardized PulseNet protocol for *E. coli* O157:H7 (27). Plugs were digested using 50 U of XbaI enzyme (New England BioLabs, Ipswich, MA). Gels were photographed using a Gel Doc 1000 (Bio-Rad), and resulting PFGE patterns were analyzed using BioNumerics software (version 6.0; Applied Maths, Kortrijk, Belgium) as described by Ju et al. (28). The PFGE pattern analyses and genetic similarity coefficients were calculated using Pearson correlation at 2.0% and a band position tolerance of 2.0%, respectively. The dendrogram was constructed using unweighted-pair group means with arithmetic averages (UPGMA).

RESULTS

In this study, we confirmed the serotypes of O157:H7 and other H types, identifying major virulence genes associated with prototypic virulent *E. coli* O157:H7, in order to determine the pathogenic potential of a collection of historic O157 isolates. Our results establish a better understanding of diversity within the O157 serogroup, particularly in the 1990s, when O157:H7 outbreaks became more frequent (17).

Strain characterization. The majority of isolates were confirmed to be serogroup O157 (380; 95%), and more than half of these O157-positive isolates carried the H7 *fliC* gene (218; 55%; *E. coli* O157:H7), 162 (41%) were O157:non-H7 strains, and 20 (5%) were non-O157:non-H7 strains (Table 2). Non-O157:non-H7 serotypes were not analyzed further in this study. The O157:non-H7 group consisted of 69 (43%) untypeable and 93 (57%) typeable H types; the majority were serotype O157:H43 (60; 65%). Many of these O157 serotypes had not been reported previously or analyzed for their pathogenic potential. The ability to ferment sorbitol was least common among O157:H7 isolates (27%), whereas nearly two-thirds (65%) of O157:non-H7 isolates were able to ferment sorbitol.

Prevalence of virulence markers. The prevalence of virulence markers differed significantly among O157:H7 and O157:non-H7 isolates (see Table S1 in the supplemental material). The majority of O157:H7 isolates carried the virulence factors for which we tested: 62% carried *stx*₁, 88% carried *stx*₂, 56% carried both *stx*₁ and *stx*₂, and 6% carried no *stx* genes. Markers and genes characteristic of O157:H7 (*eae*, *uidA*, *tir*, *terD*, and *iha*) were present at between 94 and 96% in O157:H7 isolates. Most O157:H7 isolates (96%) contained a virulence plasmid and the associated virulence genes: *hlyA* (96%), *toxB* (96%), *espP* (94%), and *katP* (94%).

None of the O157:non-H7 isolates carried *stx* genes. Although *eae* and *tir* also were absent, the presence of *terD* (19%), *iha* (31%), and *uidA* (1%) were confirmed in some of the isolates. EHEC plasmid-associated virulence genes were rare among the O157:non-H7 isolates (3%) analyzed; however, we found a higher prevalence among several non-O157:non-H7 isolates (20%). Although O157:non-H7 isolates contained *hlyA*, other EHEC plasmid-associated virulence genes (*katP*, *espP*, and *toxB*) were less frequently detected. Isolates carrying a virulence plasmid belonged mainly to O157:H2 and some O157:H43 isolates.

PFGE analysis. Using PFGE, the *E. coli* O157:H7 isolates were further characterized, and the majority of these isolates were sep-

TABLE 1 Primer sequences used in this study

Target gene	Primer sequence (5'–3')	Amplicon size (bp)	Reference
Genome located			
<i>per-rfbE</i> (O157)	CTACAGGTGAAGGTGGAATGG ATTCTCTCTTTCCTCTGCGG	327	19
<i>fliC</i> (H7)	TACCATCGCAAAAGCAACTCC GTCGGCAACGTTAGTGATACC	247	19
<i>stx</i> ₁	ATAAATCGCCATTGTTGACTAC AGAACGCCCACTGAGATCATC	180	20
<i>stx</i> ₂	GGCACTGTCTGAACTGCTCC TCGCCAGTTATCTGACATTCTG	255	20
<i>eae</i>	GACCCGGCACAAGCATAAGC CCACCTGCAGCAACAAGAGG	384	20
<i>hlyA</i>	GCATCATCAAGCGTACGTTCC AATGAGCCAAGCTGGTTAAGCT	534	20
<i>tir</i>	TCTGTTTCTGAGAATATGGGGAATA TAAAAGTTTCTGAGATCTTGATGACAT	479	21
<i>terD</i>	AGTAAAGCAGCTCCGTCAT CCGAACAGCATGGCAGTCT	434	22
<i>iha</i>	CAGTTCAGTTTCGCATTACCC GTATGGCTCTGATGCGATG	1,305	23
<i>uidA</i>	GCGAAACTGTGGAATTGGG TGATGCTCCATCACTTCTCTG	252	24
Plasmid located			
<i>espP</i>	TGTTCTGCTATGATAACCA AGTTCCGCCATTTTTCGC	642	9
<i>katP</i>	CTTCTGTTCTGATTCTTCTGG AACTTATTTCTCGCATCATCC	2,125	25
<i>toxB</i>	ATACCTACCTGCTCTGGATTGA TTCTTACCTGATCTGATGCAGC	602	26

arated from O157:non-H7, with a total of nine main clusters observed (I to IX) (Fig. 1 and 2; also see Fig. S1 in the supplemental material). Reference strains from different clonal complexes were added to the panel to determine how historical isolates clustered with those well-characterized strains. The majority of the historical O157:H7 isolates grouped in cluster II together with Sakai and other outbreak O157:H7 strains that are members of CC A6 (ESC0516, ESC0517, ESC0041, and ESC0436) (Fig. 2).

Cluster I was composed of 13 *E. coli* O157 isolates and further divided into three subclusters, Ia to Ic, with each subcluster displaying different H types (Fig. 2). Subcluster Ia was highly homogeneous (85% similarity among these PFGE patterns) and contained five O157:H2 isolates that lacked *stx*₁ and *stx*₂ but carried *iha* and *terD* and a virulence plasmid (*hlyA*), with the exception of 3821-1, which was negative for *hlyA*. Cluster Ib included isolates belong-

ing to CC A4, together with LSU-61 (O157:H7; isolated in the United States in 2001). This cluster was distant from those of other O157:H7 isolates. Within this group, isolates 3840 and 3641 lacked *stx*₁ and plasmid-encoded virulence genes (*katP*, *espP*, and *toxB*), but isolate 3641 did carry the virulence genes *uidA*, *terD*, *iha*, and *tir*. Other O157:H7 isolates differed in the presence of virulence genes and their virulence plasmids. Isolate 3840 carried only *uidA*. Isolate 3693 tested positive for all factors and carried plasmid pO157, which typically is present in CC A6 strains. Cluster Ic was composed of three O157:H11 isolates that lacked the virulence genes tested in this study.

Cluster II grouped outbreak strains from two different years together (ESC0516, ESC0517, and ESC0041 from 1983 and ESC0436 from 2006). Several of the historic isolates exhibited PFGE patterns highly similar to those of the outbreak strains and carried all of the

TABLE 2 Prevalence of virulence genes located on genomes and plasmids in isolates analyzed in this study

Serotype (400 strains total)	SF	Prevalence (%) of virulence markers from ^a :							Plasmid			
		Genome							hlyA			
		<i>stx</i> ₁	<i>stx</i> ₂	<i>eae</i>	<i>tir</i>	<i>uidA</i>	<i>terD</i>	<i>iha</i>	<i>hlyA</i>	<i>espP</i>	<i>katP</i>	<i>toxB</i>
O157:H7 (218; 55%)	27	62	98	94	96	96	96	96	96	94	94	96
O157:non-H7 (162; 41%)	66						19	31	3	1		1
Non-O157:non-H7 (20; 5%)	45	10	10	10			25	5	20	5		

^a EHEC and EAEC virulence markers include SF (sorbitol fermentation), *stx*₁ (Shiga toxin 1), *stx*₂ (Shiga toxin 2), *eae* (γ-intimin), *uidA* (β-D-glucuronidase), *tir* (putative translocated intimin receptor protein), *terD* (tellurite resistance gene), *iha* (adhesin), *hlyA* (enterohemolysin gene), *espP* (extracellular serine protease), *katP* (EHEC-catalase/peroxidase), and *toxB* (pO157 adhesion).

virulence factors found in O157:H7; however, some appeared to lack *stx*₁ (subcluster IIb). A strain isolated from beef in Malaysia in 1998, O rough:H7 MA6, also lacked *stx*₁ and clustered closely with subcluster IIb. Multiple isolates showed PFGE patterns similar to those of Sakai and ESC0516 (subcluster IIa), carrying various virulence factors and both *stx*₁ and *stx*₂ (Fig. 2).

The ability to ferment sorbitol did not appear to be restricted to any group of isolates and varied even between isolates with similar or the same PFGE patterns (3593, 3661, and 3915). The isolates of subcluster IIa were more likely to carry *stx*₁ and *stx*₂ (both 80%) than those of subcluster IIb, with *stx*₁ being present in 33% and *stx*₂ in 66% of the isolates. The lack of virulence markers (*eae*, *iha*, *terD*, *hlyA*, *espP*, and *katP*) also was independent of clustering. One isolate (3875), although belonging to cluster II, had a unique PFGE pattern.

We divided O157:non-H7 isolates into clusters by H type. Cluster III included isolates solely from pig sources; subcluster IIIa was characterized as H nontypeable (NT), but its flagellar *fliC* gene is, by sequence, most similar to the one described for *E. coli* strain A 18d (GenBank accession number AY250002.1). Virulence marker genes *iha* and *terD* appeared more frequently: 3 strains out of 4 carried *terD*, and 2 out of 4 carried *iha* genes (see Table S1 in the supplemental material). Cluster IIIb was comprised of isolates carrying H43 flagella, with 2 out of those 9 strains being *terD* positive. The two strains of cluster IV, which is more distantly related to cluster III, are serotype O157:H16 and carried none of the virulence genes analyzed in this study.

Cluster V and VI group together more closely; isolates of serotype O157:H1 (V) and O157:H42 (VI) were void of the virulence factors analyzed in this study, with the exception of isolate 3579, which carried *terD*. All H1 strains were sorbitol fermenters, while H42 isolates typically lack this ability. Two H42 isolates, 3839 and 3867, had nearly identical PFGE patterns and were negative for all of the markers tested. However, H42 isolate 3579 did exhibit sorbitol fermentation and carried *terD*.

Cluster VII contained two subclusters: VIIa and VIIb. Of the subcluster VIIa isolates, which have been isolated only from humans, all but one (3516) carried *iha* and *terD*. Both subclusters exhibited NT *fliC* genes and have been characterized as nonmotile (data not shown). The sequenced *fliC* gene of VIIa showed high sequence similarity to *E. coli* strain A 18d. The *fliC* gene of cluster VIIb strains appeared most similar, by sequence, to the one found in O121:NT.

Cluster VIII O157:H7 isolates are more similar in PFGE patterns to O55:H7 and O157:non-H7 isolates than to other O157:H7 isolates. Subcluster VIIIb, containing O157:H7 strains G5101 (isolated in the United States in 1995) and 1659, belongs to CC A7. They group closely with known ancestral O55:H7 strains (DEC5A, USDA5905, and TB182A) of subcluster IIIa. One isolate from this study, although lacking *stx*₁ (3517), is part of cluster VIIIb and carried pO157_2, which is typical of CC A7. Although none of the other O157 isolates from this study cluster closely with

O55 strains, A7 CC strains do, which indicates their closer evolutionary relationship, as suggested by other researchers (7, 29).

Finally, cluster IX contains strains carrying *iha* or *terD* and a *fliC* gene similar to that of CC clusters IIIa and VIIa (*E. coli* strain A 18d).

DISCUSSION

The analysis of this comprehensive collection of historic *E. coli* isolates provided a unique opportunity to examine the diversity and pathogenicity of the O157 serogroup. O157:H7 isolates in this study carried most genes previously associated with virulence and cluster together in PFGE cluster II, comprising strains of CC A6. Isolates were highly diverse, suggesting recent genomic divergences among isolates, similar to what has been shown previously for the O157:H7 serotype (7, 30).

Many of the isolates in this study were sorbitol fermenters, which is not typical for classical *E. coli* O157:H7. Thus far, the CC A4 SF O157 nonmotile variant has been found in Germany and other European countries, and only rarely has it been documented in North America (31, 32). Typical NSF O157:H7 strains are part of CC A6, carrying *stx*₁ and *stx*₂, and are found worldwide (7). The majority SF O157:H7 isolates found in this study cluster in CC A6 and carry various virulence genes, including *stx*. Because *stx* is carried on phages, it can be gained or lost easily in isolates; therefore, *stx* is rarely used as an indicator of relatedness or, particularly for *stx*₁, virulence (33). Another SF O157:H7 cluster, CC A7, comprises isolates that are genetically distinct and not ancestral to CC A6 isolates (9). Mainly isolated in Japan and the United States, these strains carry *stx*₁ and *stx*₂ and show the unique presence of virulence plasmid pO157_2, which distinguishes this group from CC A4 SF O157:H carrying pSFO157 and A6 CC isolates carrying pO157 (9, 34).

Virulent serotypes of O157 that exhibit the ability to ferment sorbitol lead to additional challenges for detection. As the NSF trait is commonly used in culture-based identification methods, those techniques will be unable to detect these pathogens (8, 35). These findings underline the importance of the presence of SF O157:H7 strains that are capable of causing disease; however, they are more difficult to quickly detect in routine analysis. Detection challenges are arising already due to the emergence of SF non-O157 STEC serogroups, since, without suitable differential agar media, a phenotypic discrimination is problematic (36). This highlights the importance of clearly distinguishing among O157:H7 strains for the timely identification of cases and the prevention of disease, as well as the need to screen for *E. coli* O157:H7 on sorbitol MacConkey agar, since the ability to ferment sorbitol appears to be characteristic for this serotype (35, 37).

Interestingly, only isolate 3517 in this study clusters closely with the CC A7 strains G5101 and 1659 (both ST65), carrying pO157_2. Previous research showed that, considering their plasmids, O157:H7 strains likely evolved into separate lineages (9). The CC A7 group is more separate from other CC A6 O157:H7

FIG 1 Cluster I and III to IX O157:H7 isolate PFGE results. EHEC virulence markers, located on the chromosome and plasmid, and strain identification numbers are shown. The dendrogram was created by the unweighted-pair group method using average linkages and was generated with BioNumerics software (Applied Maths, St-Martens-Latem, Belgium) by using a Pearson correction with a 2.0% lane optimization and 2.5% band position tolerance. The scale above the dendrogram indicates percent similarity. Virulence gene markers are the following: SF, sorbitol-fermentation ability; *eaeA*, γ -intimin; *hlyA*, enterohemolysin; *stx*₁, Shiga toxin 1; *stx*₂, Shiga toxin 2; O157, presence of the *per* gene, specific for O157 antigen; H7, presence of the *fliC* (H7) gene, specific for H7 antigen; *iha*, adhesin; *terD*, tellurite resistance; *uidA*, β -D-glucuronidase; *tir*, putative translocated intimin receptor protein; *tox*B, pO157 adhesion; *espP*, extracellular serine protease; *katP*, EHEC-catalase/peroxidase. Strain numbers, cluster numbers (I to XI, except II) and subclusters are indicated. Results for cluster II are shown in Fig. 2.

isolates, instead clustering closely with O55 strains, indicating a closer relatedness to this ancestral O55:H7 group. A definitive evolutionary pathway remains unclear. Further analysis of atypical isolates could help reveal a more accurate model of how highly pathogenic O157:H7 strains have evolved. Isolates of this group have been shown to carry identical traits (including the lack of *katP* and *espP*) and share ST65 with O157:H7 strains from Japan isolated in 1996 to 1997 (7, 38). Most notably, those strains appear to cause mild or no symptoms in patients and thus far have been implicated only in sporadic cases of illness (38). The distinct separation of these serotypes from typical A6 strains, the few numbers of isolates found in this study (1 of 218 isolates), and the epidemiology of CC A7 isolates suggest that members of CC A7 are less prevalent in the United States, possibly worldwide. They might have a lower virulence potential than other O157:H7 strains, which might explain why they have not yet been associated with major outbreaks.

It was demonstrated previously that serotypes with the same H types cluster closer together, as was seen for O26:H11 and O111:H11, which are more related to each other than to O111:H11 or O111:H8 (39). In our study, CC A4 strains clustered closely together with H2 and H11 isolates, whereas other H types are clearly distinct from typical O157:H7. Until now, the non-O157 STEC serotypes most commonly implicated in illnesses have been O26:H11 (or NM), O111:H8, O111:H11 (or NM), O103:H2, O103:H11, O103:H25 (or NM), O121:H19, O121:H7, O45:H2 (or NM), and O145:NM (40, 41). Therefore, the most common H types are H2 and H11, both of which are present in different O types (O103 and O45 for H2 and O26, O111, and O103 for H11). Therefore, the finding in this study that O157:H2, O157:NM (H7), and O157:H11 group so closely together indicates another path for the emergence of virulent STEC serotypes. Notably, O157:H2 and O157:H11 also are the H types in this study which group closer together than other H types. Since these H types also have been found in virulent non-O157 STEC, it is likely there is some connection between the H types H2 and H11 and the emergence of more virulent STEC serotypes.

The H2 group isolates all carry *hlyA*, *iha*, and *terD* virulence genes typical of O157:H7. This indicates that sharing the same niche increases the likelihood of horizontal transfer, enabling the integration of virulence factors in genomes, which then might be subjected to selection pressures favoring isolates with pathogenic potential. Further analysis of O157:H2 and O157:H11 and their phylogenetic relatedness to other H2 and H11 isolates could help shed light on the emergence of pathogenic non-O157 isolates.

Generally, O157:non-H7 strains of *E. coli* do not carry virulence genes and, to date, have rarely been implicated in disease (12). In this study, we found several non-H7 isolates clustering closely together according to H type and carrying important STEC virulence genes. Although our analyses found no *stx*₁ or *stx*₂, *eae*, and *uidA* in the current collection of 400 isolates, *iha* and *terD* have been confirmed in other isolates (42). As phages and O islands can be horizontally transferred, it is conceivable that *iha* and *terD* genes could be laterally transferred from existing pathogenic groups to other *E. coli* strains (43). For example, Ju et al. demonstrated that O-43/48-associated virulence genes are frequently present in *eae*-positive STEC, whereas *eae*-negative strains typically carry only *iha* (44). The prevalence of *iha* among serotypes implicated in severe disease as well as serotypes that are rarely associated with disease indicates that the presence of *iha* is

not directly related to severe outcomes (44). However, since most isolates from humans (clusters VII and IX) appear more likely to carry *iha* than the other isolates in our study, we should not exclude the possibility of disease complications attributable to this virulence factor. Similar cautions apply to assessing the importance of markers for tellurium resistance (*terD*) (45). The presence of such genes in serotypes that share similar niches but have not yet been implicated in illness (a pattern also seen among O157:H2 isolates [cluster Ia] and other pig isolates [cluster II and VIIa]) indicates that these factors alone are not indicators of pathogenicity but again point out that virulence factors are frequently being acquired by other, currently nonpathogenic, serotypes.

The majority of isolates in this study were devoid of virulence plasmids (and none carried the other pO157 typical virulence genes, *katP*, *espP*, and *toxB*). The majority of isolates carrying a virulence plasmid containing *hlyA* belong to serotype O157:H2 and mostly were isolated from rabbit and pig. As stated above, this H type is more closely related to O157:H7/H than other H types, which usually carry an EHEC plasmid that is partially responsible for its high pathogenicity. Other flagellar H2 STEC isolates (O103:H2 and O45:H2) also were shown to carry EHEC virulence plasmids and, due to previous findings of a close phylogenetic relationship of H types to O types, the presence of plasmids in this serotype could be a further indicator of a close relatedness to other H2 types (39, 46, 47). Therefore, the O157:H2 serogroup should be investigated further for its pathogenic potential and the possible emergence of another important serotype.

In conclusion, the highly diverse O157 serogroup can be divided into corresponding H types, and only serotype O157:H7 and closely related serotypes carry the typical virulence markers of highly pathogenic EHEC. Even when non-H7 isolates were void of other virulence markers, *iha* and *terD* were found frequently in almost all isolates regardless of their H type. Two particular H types, H2 and H11, clustered more closely to O157:H7 and carried virulence genes found in the most commonly found non-O157 H types associated with severe disease (O26:H11, O111:H11, O103:H2, O103:H11, and O45:H2). By characterizing collections of historic isolates and their corresponding plasmids, we can better understand both the evolution of known pathogens, such as O157:H7, and the factors that could lead to new, highly pathogenic *E. coli* serotypes.

REFERENCES

1. Blanco JE, Blanco M, Alonso MP, Mora A, Dahbi G, Coira MA, Blanco J. 2004. Serotypes, virulence genes, and intimin types of Shiga toxin (verotoxin)-producing *Escherichia coli* isolates from human patients: prevalence in Lugo, Spain, from 1992 through 1999. *J Clin Microbiol* 42:311–319. <http://dx.doi.org/10.1128/JCM.42.1.311-319.2004>.
2. Nataro JP, Kaper JB. 1998. Diarrheagenic *Escherichia coli*. *Clin Microbiol Rev* 11:142–201.
3. Gould LH, Mody RK, Ong KL, Clogher P, Cronquist AB, Garman KN, Lathrop S, Medus C, Spina NL, Webb TH, White PL, Wymore K, Gierke RE, Mahon BE, Griffin PM. 2013. Increased recognition of non-O157 Shiga toxin-producing *Escherichia coli* infections in the United States during 2000–2010: epidemiologic features and comparison with *E. coli* O157 infections. *Foodborne Pathog Dis* 10:453–460. <http://dx.doi.org/10.1089/fpd.2012.1401>.
4. Lim JY, Yoon J, Hovde CJ. 2010. A brief overview of *Escherichia coli* O157:H7 and its plasmid O157. *J Microbiol Biotechnol* 20:5–14.
5. Bielaszewska M, Mellmann A, Zhang W, Kock R, Fruth A, Bauwens A, Peters G, Karch H. 2011. Characterisation of the *Escherichia coli* strain associated with an outbreak of haemolytic uraemic syndrome in Germany, 2011: a microbiological study. *Lancet Infect Dis* 11:671–676. [http://dx.doi.org/10.1016/S1473-3099\(11\)70165-7](http://dx.doi.org/10.1016/S1473-3099(11)70165-7).
6. Eppinger M, Mammel MK, Lederer JE, Ravel J, Cebula TA. 2011. Genomic

- anatomy of *Escherichia coli* O157:H7 outbreaks. *Proc Natl Acad Sci U S A* 108:20142–20147. <http://dx.doi.org/10.1073/pnas.1107176108>.
7. Feng PC, Monday SR, Lacher DW, Allison L, Siitonen A, Keys C, Eklund M, Nagano H, Karch H, Keen J, Whittam TS. 2007. Genetic diversity among clonal lineages within *Escherichia coli* O157:H7 stepwise evolutionary model. *Emerg Infect Dis* 13:1701–1706. <http://dx.doi.org/10.3201/eid1311.070381>.
 8. Feng P. 1995. *Escherichia coli* serotype O157:H7: novel vehicles of infection and emergence of phenotypic variants. *Emerg Infect Dis* 1:47–52. <http://dx.doi.org/10.3201/eid0102.950202>.
 9. Rump LV, Meng J, Strain EA, Cao G, Allard MW, Gonzalez-Escalona N. 2012. Complete DNA sequence analysis of enterohemorrhagic *Escherichia coli* plasmid pO157_2 in beta-glucuronidase-positive *E. coli* O157:H7 reveals a novel evolutionary path. *J Bacteriol* 194:3457–3463. <http://dx.doi.org/10.1128/JB.00197-12>.
 10. Jenke C, Harmsen D, Weniger T, Rothganger J, Hyytia-Trees E, Bielewska M, Karch H, Mellmann A. 2010. Phylogenetic analysis of enterohemorrhagic *Escherichia coli* O157, Germany, 1987–2008. *Emerg Infect Dis* 16:610–616. <http://dx.doi.org/10.3201/eid1604.091361>.
 11. Tatsuno I, Horie M, Abe H, Miki T, Makino K, Shinagawa H, Taguchi H, Kamiya S, Hayashi T, Sasakawa C. 2001. *tox*B gene on pO157 of enterohemorrhagic *Escherichia coli* O157:H7 is required for full epithelial cell adherence phenotype. *Infect Immun* 69:6660–6669. <http://dx.doi.org/10.1128/IAI.69.11.6660-6669.2001>.
 12. Feng PC, Keys C, Lacher D, Monday SR, Shelton D, Rozand C, Rivas M, Whittam T. 2010. Prevalence, characterization and clonal analysis of *Escherichia coli* O157: non-H7 serotypes that carry *eae* alleles. *FEMS Microbiol Lett* 308:62–67. <http://dx.doi.org/10.1111/j.1574-6968.2010.01990.x>.
 13. Makino S, Asakura H, Shirahata T, Ikeda T, Takeshi K, Arai K, Nagasawa M, Abe T, Sadamoto T. 1999. Molecular epidemiological study of a mass outbreak caused by enteropathogenic *Escherichia coli* O157:H45. *Microbiol Immunol* 43:381–384. <http://dx.doi.org/10.1111/j.1348-0421.1999.tb02419.x>.
 14. Bell BP, Goldoft M, Griffin PM, Davis MA, Gordon DC, Tarr PI, Bartleson CA, Lewis JH, Barrett TJ, Wells JG. 1994. A multistate outbreak of *Escherichia coli* O157:H7-associated bloody diarrhea and hemolytic uremic syndrome from hamburgers. The Washington experience. *JAMA* 272:1349–1353. <http://dx.doi.org/10.1001/jama.1994.03520170059036>.
 15. Tuttle J, Gomez T, Doyle MP, Wells JG, Zhao T, Tauxe RV, Griffin PM. 1999. Lessons from a large outbreak of *Escherichia coli* O157:H7 infections: insights into the infectious dose and method of widespread contamination of hamburger patties. *Epidemiol Infect* 122:185–192. <http://dx.doi.org/10.1017/S0950268898001976>.
 16. Yukioka H, Kurita S. 1997. *Escherichia coli* O157 infection disaster in Japan, 1996. *Eur J Emerg Med* 4:165. <http://dx.doi.org/10.1097/00063110-199709000-00009>.
 17. Rangel JM, Sparling PH, Crowe C, Griffin PM, Swerdlow DL. 2005. Epidemiology of *Escherichia coli* O157:H7 outbreaks, United States, 1982–2002. *Emerg Infect Dis* 11:603–609. <http://dx.doi.org/10.3201/eid1104.040739>.
 18. Lacher DW, Steinsland H, Blank TE, Sonnenberg MS, Whittam TS. 2007. Molecular evolution of typical enteropathogenic *Escherichia coli*: clonal analysis by multilocus sequence typing and virulence gene allelic profiling. *J Bacteriol* 189:342–350. <http://dx.doi.org/10.1128/JB.01472-06>.
 19. Wang J, Qiu M, Lin X. 2002. Development of a rapid separating incubation method for detection of *Escherichia coli* O157:H7 in food. *Wei Sheng Yan Jiu* 31:44–46.
 20. Paton AW, Paton JC. 1998. Detection and characterization of Shiga toxin-producing *Escherichia coli* by using multiplex PCR assays for *stx*1, *stx*2, *eae*A, enterohemorrhagic *E. coli* *hly*A, *rfb*O111, and *rfb*O157. *J Clin Microbiol* 36:598–602.
 21. Ogura Y, Ooka T, Asadulghani Terajima J, Nougayrede JP, Kurokawa K, Tashiro K, Tobe T, Nakayama K, Kuhara S, Oswald E, Watanabe H, Hayashi T. 2007. Extensive genomic diversity and selective conservation of virulence-determinants in enterohemorrhagic *Escherichia coli* strains of O157 and non-O157 serotypes. *Genome Biol* 8:R138. <http://dx.doi.org/10.1186/gb-2007-8-7-r138>.
 22. Bielewska M, Tarr PI, Karch H, Zhang W, Mathys W. 2005. Phenotypic and molecular analysis of tellurite resistance among enterohemorrhagic *Escherichia coli* O157:H7 and sorbitol-fermenting O157:NM clinical isolates. *J Clin Microbiol* 43:452–454. <http://dx.doi.org/10.1128/JCM.43.1.452-454.2005>.
 23. Schmidt H, Zhang WL, Hemmrich U, Jelacic S, Brunder W, Tarr PI, Dobrindt U, Hacker J, Karch H. 2001. Identification and characterization of a novel genomic island integrated at *selC* in locus of enterocyte effacement-negative, Shiga toxin-producing *Escherichia coli*. *Infect Immun* 69:6863–6873. <http://dx.doi.org/10.1128/IAI.69.11.6863-6873.2001>.
 24. Cebula TA, Payne WL, Feng P. 1995. Simultaneous identification of strains of *Escherichia coli* serotype O157:H7 and their Shiga-like toxin type by mismatch amplification mutation assay-multiplex PCR. *J Clin Microbiol* 33:248–250.
 25. Brunder W, Schmidt H, Karch H. 1996. *KatP*, a novel catalase-peroxidase encoded by the large plasmid of enterohemorrhagic *Escherichia coli* O157:H7. *Microbiology* 142(Part 11):3305–3315. <http://dx.doi.org/10.1099/13500872-142-11-3305>.
 26. Tarr CL, Large TM, Moeller CL, Lacher DW, Tarr PI, Acheson DW, Whittam TS. 2002. Molecular characterization of a serotype O121:H19 clone, a distinct Shiga toxin-producing clone of pathogenic *Escherichia coli*. *Infect Immun* 70:6853–6859. <http://dx.doi.org/10.1128/IAI.70.12.6853-6859.2002>.
 27. Ribot EM, Fair MA, Gautam 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. <http://dx.doi.org/10.1089/fpd.2006.3.59>.
 28. Ju W, Shen J, Li Y, Toro MA, Zhao S, Ayers S, Najjar MB, Meng J. 2012. Non-O157 Shiga toxin-producing *Escherichia coli* O157:H7 in retail ground beef and pork in the Washington D.C. area. *Food Microbiol* 32:371–377. <http://dx.doi.org/10.1016/j.fm.2012.07.017>.
 29. Zhou Z, Li X, Liu B, Beutin L, Xu J, Ren Y, Feng L, Lan R, Reeves PR, Wang L. 2010. Derivation of *Escherichia coli* O157:H7 from its O55:H7 precursor. *PLoS One* 5:e8700. <http://dx.doi.org/10.1371/journal.pone.0008700>.
 30. Ogura Y, Kurokawa K, Ooka T, Tashiro K, Tobe T, Ohnishi M, Nakayama K, Morimoto T, Terajima J, Watanabe H, Kuhara S, Hayashi T. 2006. Complexity of the genomic diversity in enterohemorrhagic *Escherichia coli* O157 revealed by the combinational use of the O157 Sakai OligoDNA microarray and the whole genome PCR scanning. *DNA Res* 13:3–14. <http://dx.doi.org/10.1093/dnares/dsi026>.
 31. Feng P, Lampel KA, Karch H, Whittam TS. 1998. Genotypic and phenotypic changes in the emergence of *Escherichia coli* O157:H7. *J Infect Dis* 177:1750–1753. <http://dx.doi.org/10.1086/517438>.
 32. Karch H, Bohm H, Schmidt H, Gunzer F, Aleksic S, Heesemann J. 1993. Clonal structure and pathogenicity of Shiga-like toxin-producing, sorbitol-fermenting *Escherichia coli* O157:H. *J Clin Microbiol* 31:1200–1205.
 33. Asadulghani M, Ogura Y, Ooka T, Itoh T, Sawaguchi A, Iguchi A, Nakayama K, Hayashi T. 2009. The defective prophage pool of *Escherichia coli* O157: prophage-prophage interactions potentiate horizontal transfer of virulence determinants. *PLoS Pathog* 5:e1000408. <http://dx.doi.org/10.1371/journal.ppat.1000408>.
 34. Brunder W, Karch H, Schmidt H. 2006. Complete sequence of the large virulence plasmid pSFO157 of the sorbitol-fermenting enterohemorrhagic *Escherichia coli* O157:H- strain 3072/96. *Int J Med Microbiol* 296:467–474. <http://dx.doi.org/10.1016/j.ijmm.2006.05.005>.
 35. Werber D, Bielewska M, Frank C, Stark K, Karch H. 2011. Watch out for the even eviler cousin-sorbitol-fermenting *E. coli* O157. *Lancet* 377:298–299. [http://dx.doi.org/10.1016/S0140-6736\(11\)60090-1](http://dx.doi.org/10.1016/S0140-6736(11)60090-1).
 36. Medina MB, Sheller WL, Fratamico PM, Fortis L, Tillman G, Narang N, Cray WC, Jr, Esteban E, Debroy A. 2012. Latex agglutination assays for detection of non-O157 Shiga toxin-producing *Escherichia coli* serogroups O26, O45, O103, O111, O121, and O145. *J Food Prot* 75:819–826. <http://dx.doi.org/10.4315/0362-028X.JFP-11-430>.
 37. Diaz S, Vidal D, Herrera-Leon S, Sanchez S. 2011. Sorbitol-fermenting, beta-glucuronidase-positive, Shiga toxin-negative *Escherichia coli* O157:H7 in free-ranging red deer in south-central Spain. *Foodborne Pathog Dis* 8:1313–1315. <http://dx.doi.org/10.1089/fpd.2011.0923>.
 38. Nagano H, Okui T, Fujiwara O, Uchiyama Y, Tamate N, Kumada H, Morimoto Y, Yano S. 2002. Clonal structure of Shiga toxin (Stx)-producing and beta-D-glucuronidase-positive *Escherichia coli* O157:H7 strains isolated from outbreaks and sporadic cases in Hokkaido, Japan. *J Med Microbiol* 51:405–416.
 39. Ju W, Cao G, Rump L, Strain E, Luo Y, Timme R, Allard M, Zhao S, Brown E, Meng J. 2012. Phylogenetic analysis of non-O157 Shiga toxin-producing *Escherichia coli* by whole genome sequencing. *J Clin Microbiol* 50:4123–4127. <http://dx.doi.org/10.1128/JCM.02262-12>.

40. 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>.
41. Wickham ME, Lupp C, Mascarenhas M, Vazquez A, Coombes BK, Brown NF, Coburn BA, Deng W, Puente JL, Karmali MA, Finlay BB. 2006. Bacterial genetic determinants of non-O157 STEC outbreaks and hemolytic-uremic syndrome after infection. *J Infect Dis* **194**:819-827. <http://dx.doi.org/10.1086/506620>.
42. Taylor DE, Rooker M, Keelan M, Ng LK, Martin I, Perna NT, Burland NT, Blattner FR. 2002. Genomic variability of O islands encoding tellurite resistance in enterohemorrhagic *Escherichia coli* O157:H7 isolates. *J Bacteriol* **184**: 4690-4698. <http://dx.doi.org/10.1128/JB.184.17.4690-4698.2002>.
43. Castillo A, Eguiarte LE, Souza V. 2005. A genomic population genetics analysis of the pathogenic enterocyte effacement island in *Escherichia coli*: the search for the unit of selection. *Proc Natl Acad Sci U S A* **102**:1542-1547. <http://dx.doi.org/10.1073/pnas.0408633102>.
44. Ju W, Shen J, Toro M, Zhao S, Meng J. 2013. Distribution of pathogenicity islands OI-122, OI-43/48, and OI-57 and a high-pathogenicity island in Shiga toxin-producing *Escherichia coli*. *Appl Environ Microbiol* **79**:3406-3412. <http://dx.doi.org/10.1128/AEM.03661-12>.
45. Taylor DE. 1999. Bacterial tellurite resistance. *Trends Microbiol* **7**:111-115. [http://dx.doi.org/10.1016/S0966-842X\(99\)01454-7](http://dx.doi.org/10.1016/S0966-842X(99)01454-7).
46. Fratamico PM, Sackitey SK, Wiedmann M, Deng MY. 1995. Detection of *Escherichia coli* O157:H7 by multiplex PCR. *J Clin Microbiol* **33**:2188-2191.
47. Toro M, Cao G, Ju W, Allard M, Barrangou R, Zhao S, Brown E, Meng J. 2014. Association of clustered regularly interspaced short palindromic repeat (CRISPR) elements with specific serotypes and virulence potential of Shiga toxin-producing *Escherichia coli*. *Appl Environ Microbiol* **80**: 1411-1420. <http://dx.doi.org/10.1128/AEM.03018-13>.