JCM
Journals.ASM.org

Phylogenetic Analysis of Non-O157 Shiga Toxin-Producing *Escherichia coli* Strains by Whole-Genome Sequencing

Wenting Ju,[a] Guojie Cao,[a] Lydia Rump,[a] Errol Strain,[b] Yan Luo,[b] Ruth Timme,[c] Marc Allard,[c] Shaohua Zhao,[d] Eric Brown,[c] and Jianghong Meng[a]

Department of Nutrition and Food Science and Joint Institute for Food Safety and Applied Nutrition, University of Maryland, College Park, Maryland, USA[a]; Branch of Biostatistics, Center for Food Safety and Applied Nutrition, Food and Drug Administration, College Park, Maryland, USA[b]; Division of Microbiology, Center for Food Safety and Applied Nutrition, Food and Drug Administration, College Park, Maryland, USA[c]; and Division of Animal and Food Microbiology, Office of Research, Center for Veterinary Medicine, Food and Drug Administration, Laurel, Maryland, USA[d]

Non-O157 Shiga toxin-producing *Escherichia coli* (STEC) strains are emerging food-borne pathogens causing life-threatening diseases and food-borne outbreaks. A better understanding of their evolution provides a framework for developing tools to control food safety. We obtained 15 genomes of non-O157 STEC strains, including O26, O111, and O103 strains. Phylogenetic trees revealed a close relationship between O26:H11 and O111:H11 and a scattered distribution of O111. We hypothesize that STEC serotypes with the same H antigens might share common ancestors.

Shiga toxin-producing *Escherichia coli* (STEC) strains are deadly pathogens, causing hemorrhagic colitis (HC) and hemolytic-uremic syndrome (HUS) (4, 13). There are two categories of surface antigens (O somatic and H flagellar), whose combinations are used to classify *E. coli*. *E. coli* O157:H7 has caused more outbreaks and HUS cases in the United States than any other serotype. However, there is a growing concern about the health risk of non-O157 STEC (1), as more than 470 serotypes of STEC are associated with human diseases (2). In the United States, non-O157 STEC causes an estimated 112,752 cases of illness each year, which is more than the number of cases (estimated at 63,153) caused by *E. coli* O157:H7 (16). Among the non-O157 STEC strains, serogroups O26, O111, and O103 are considered the most clinically important and frequently identified non-O157 STEC strains in severe diseases and food-borne outbreaks (3, 14, 19). In this study, we used whole-genome sequencing data to examine the phylogenetic relationship of non-O157 STEC strains for a better understanding of the evolutionary history of these emerging pathogens.

Fifteen STEC strains representing different pulsed-field gel electrophoresis (PFGE) patterns, isolation years, hosts, and *stx* gene profiles, including O111:H11, O111:H8, O26:H11, O103:H2, and O103:H25, were selected for whole-genome sequencing analysis using the 454 pyrosequencing system (FLX; Roche, Branford, CT) to obtain draft genomes (Table 1). In addition, 28 *E. coli* published genomes were included for phylogenetic study (Table 1). The genome sizes of the 15 STEC strains ranged from 5.26 Mbp to 6.01 Mbp (Table 1). Multiple sequence alignment of all 43 genomes was performed using Mauve (5), and approximately 183,470 single nucleotide polymorphisms (SNPs) were identified.

Pulsed-field gel electrophoresis (PFGE) with XbaI was performed according to a non-O157 PulseNet protocol (http://www.pulsenetinternational.org/SiteCollectionDocuments/pfge/5%201_5%202_5%204_PNetStand_Ecoli_with_Sflexneri.pdf) and analyzed with BioNumerics software (Applied Maths, Austin, TX) using Dice coefficients and unweighted pair group means with arithmetic averages (UPGMA) to construct a dendrogram with a 1.5% band position tolerance. *eae* subtypes were determined using PCR-restriction fragment length polymorphism (RFLP) as described by Tramuta et al. (20). The 15 STEC strains were grouped into two main clusters (Fig. 1) that separated H11 strains (O111:H11 and O26:H11) from H8 strains (O111:H8). However, PFGE was not able to differentiate O111:H11 and O26:H11. In addition, the O26:H11 and O111:H11 strains shared the same *eae* subtype (β), while the O111:H8 strains contained θ. It appeared that O111:H11 and O26:H11 were more closely related to each other than either was to O111:H8, according to PFGE profiles and virulence gene-associated elements in the genomes.

Seven housekeeping genes (*aspC*, *clpX*, *fadD*, *icdA*, *lysP*, *mdh*, and *uidA*) extracted from genomes were selected for multilocus sequence typing (MLST) analysis as previously described for pathogenic *E. coli* (http://www.shigatox.net/ecmlst/protocols/index.html). The MLST analysis was performed using MEGA 5.05 (17) with 2,000 iterations (model, maximum composite likelihood; substitution, transitions plus transversions; gamma). The O111:H11, O26:H11, and O111:H8 strains formed one branch in the MLST dendrogram, with O26:H11 and O111:H11 clustering together in a lineage sister to the O111:H8 strains (Fig. 2). It is interesting that O26:H11 strain DEC10B clustered with the O111:H11 strains. Furthermore, five strains sharing the H2 antigen clustered together regardless of O serotypes. Genomic analysis revealed that O111:NM strain DEC12C carried *fliC* for the H2 gene.

To further explore evolutionary relatedness, a parsimony phylogenetic tree based on whole-genome-wide SNPs was performed with 10,000 iterations by TNT (tree analysis using new technology) (9). Similarly to data shown by PFGE and MLST, the phylogenetic tree demonstrated that O26:H11 strains belonged to the same clade as did the O111:H11 and O111:H8 strains but grouped

Received 23 August 2012 **Returned for modification** 14 September 2012
Accepted 18 September 2012
Published ahead of print 10 October 2012
Address correspondence to Jianghong Meng, jmeng@umd.edu.
W.J. and G.C. contributed equally to this work.
Copyright © 2012, American Society for Microbiology. All Rights Reserved.
doi:10.1128/JCM.02262-12

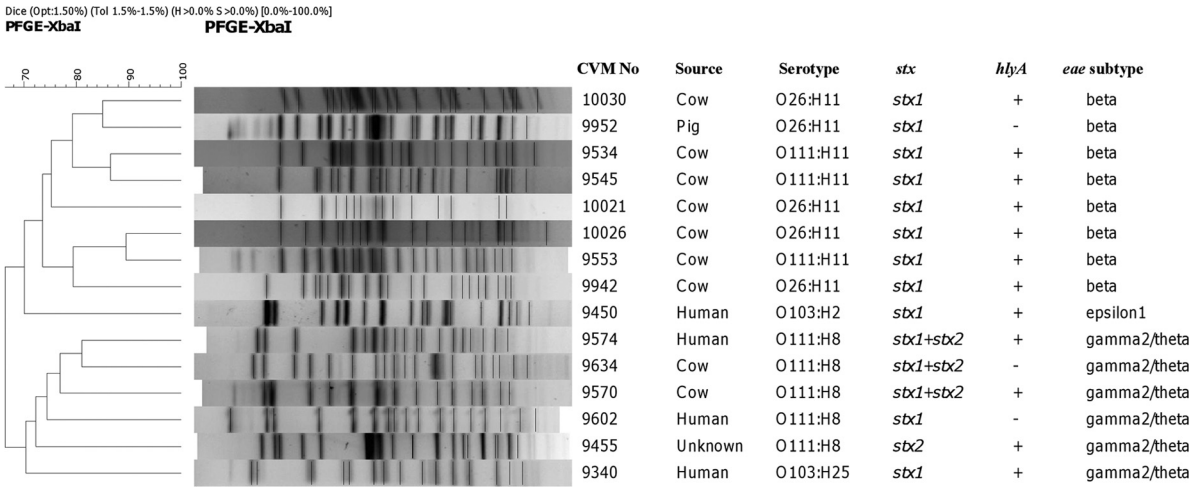
TABLE 1 Serotypes, pathotypes, toxin genotypes, sources, and genome sizes of *Escherichia coli* strains used in this study[a]

Strain	Serotype	Pathotype[b]	Shiga toxin gene	Source	Size (Mbp)	Accession no.
CVM10021	O26:H11	STEC	stx_1	Cow	5.50	AKAZ00000000
CVM9942	O26:H11	STEC	stx_1	Cow	5.62	AJVW00000000
CVM10026	O26:H11	STEC	stx_1	Cow	5.57	AJVX00000000
CVM10030	O26:H11	STEC	stx_1	Cow	5.50	AKBA00000000
CVM9952	O26:H11	STEC	stx_1	Pig	5.50	AKBC00000000
CVM9634	O111:H8	STEC	stx_1 + stx_2	Cow	5.78	AKAW00000000
CVM9602	O111:H8	STEC	stx_1	Human	5.10	AKAV00000000
CVM9574	O111:H8	STEC	stx_1 + stx_2	Human	5.36	AJVV00000000
CVM9570	O111:H8	STEC	stx_1 + stx_2	Cow	5.51	AJVU00000000
CVM9545	O111:H11	STEC	stx_1	Cow	5.61	AJVT00000000
CVM9455	O111:H11	STEC	stx_2	Unknown	6.01	AKAX00000000
CVM9534	O111:H11	STEC	stx_1	Cow	5.46	AJVS00000000
CVM9553	O111:H11	STEC	stx_1	Cow	5.60	AKAY00000000
CVM9340	O103:H25	STEC	stx_1	Human	5.26	AJVQ00000000
CVM9450	O103:H2	STEC	stx_1	Human	5.39	AJVR00000000
CFT073	O6:K2:H1	UPEC		Unknown	5.23	AE014075
Sakai	O157:H7	STEC	stx_1 + stx_2	Human	5.59	BA000007
CB9615	O55:H7	EPEC		Human	5.39	NC_013941
4865/96	O145:H28	STEC	stx_2	Human	5.23	AGTL00000000
53638	O144:?	EIEC		Unknown	5.07	AAKB00000000
101-1	O−:H10	EAEC		Human	4.98	AAMK00000000
MG1655	Unknown	Commensal		Unknown	4.64	NC_000913
5.0959	H121:H19	STEC	stx_2	Unknown	5.37	AEZX00000000
TY-2482	O104:H4	EAEC + STEC	stx_2	Human	5.29	AFOG00000000
CL-3	O113:H21	STEC	stx_2	Human	5.05	AGTH00000000
B2F1	O91:H21	STEC	stx_2	Human	5.01	AGTI00000000
E24377A	O139:H28	ETEC		Unknown	4.97	NC_009801
DEC12B	O111:H2	STEC	stx_2	Human	5.49	AIHB00000000
DEC12C	O111:NM	STEC	stx_2	Human	5.45	AIHC00000000
E22	O103:H2	EPEC		Unknown	5.53	AAJV00000000
03-EN-705	O45:H2	STEC	stx_1	Human	5.3	AGTK00000000
12009	O103:H2	STEC	stx_1 + stx_2	Human	5.45	NC_013353
E110019	O111:H9	EPEC		Human	5.38	AAJW00000000
DEC15A	O111:H21	EPEC		Human	5.25	AIHO00000000
DEC15E	O111:H21	EPEC		Human	5.23	AIHS00000000
DEC8E	O111:H8	STEC	stx_1	Human	5.32	AIGJ00000000
DEC8B	O111:H8	STEC	stx_1 + stx_2	Human	5.37	AIGG00000000
11128	O111:H−	STEC	stx_1 + stx_2	Human	5.37	NC_013364
DEC8D	O111:H11	DEC		Human	5.46	AIGI00000000
DEC8C	O111:H11	STEC	stx_1	Cow	5.91	AIGH00000000
DEC10B	O26:H11	STEC	stx_1	Human	5.58	AIGQ00000000
EPECCa14	O26:H11	STEC	stx_1	Unknown	5.44	ADUN00000000
11368	O26:H11	STEC	stx_1	Human	5.69	NC_013361

[a] Data on strains named with CVM were from this study; the rest were from GenBank.
[b] STEC, Shiga toxin-producing *Escherichia coli*; EPEC, enteropathogenic *Escherichia coli*; EIEC, enteroinvasive *Escherichia coli*; ETEC, enterotoxigenic *Escherichia coli*; EAEC, enteroaggregative *Escherichia coli*; DEC, diarrheagenic *Escherichia coli*.

more closely with strains of the same H type (O111:H11) (Fig. 3). In the TNT tree, *E. coli* O26:H11 DEC10B grouped with the O111:H11 strains as shown in the MLST dendrogram (Fig. 2), indicating a close relationship between these strains. All O111 strains but O111:H2 formed one clade (Fig. 3), including O111 enteropathogenic *E. coli* (EPEC), O111:H8, and O111:H11. Additionally, we reconstructed a maximum likelihood (ML) tree by using Garli-2.0 (22) and a Bio Neighbor Joining (BioNJ) (8) tree by using SeaView 4 (10) (data not shown), displaying similar phylogenetic relationships with the TNT tree with minor differences. For example, in the BioNJ tree, *E. coli* O26:H11 DEC10B grouped with the O26:H11 strains. The H2 strains were all closely clustered together in all phylogenetic trees, including the TNT, ML, and BioNJ trees as shown in the MLST dendrogram. The phylogenetic trees also supported the idea that STEC O113:H21 and O91:H21 were closely related (Fig. 3).

The phylogenetic trees indicated that a common ancestor might exist for strains of the same H type. For example, the H11 strains, including O26:H11 and O111:H11 (Fig. 2 and 3), shared a common ancestor; the H2 strains with different O antigens, namely, the H2 group in Fig. 3, shared a common ancestor as well. Previous studies also have suggested that STEC strains with the same H antigens might share common ancestors. For example, Iguchi et al. (11) indicated a close relatedness between STEC O103:H11 and O26:H11 shown by MLST analysis, in which the strains shared the same *eae* subtype (β-*eae*) that was also inserted



CVM No	Source	Serotype	*stx*	*hlyA*	*eae* subtype
10030	Cow	O26:H11	*stx1*	+	beta
9952	Pig	O26:H11	*stx1*	-	beta
9534	Cow	O111:H11	*stx1*	+	beta
9545	Cow	O111:H11	*stx1*	+	beta
10021	Cow	O26:H11	*stx1*	+	beta
10026	Cow	O26:H11	*stx1*	+	beta
9553	Cow	O111:H11	*stx1*	+	beta
9942	Cow	O26:H11	*stx1*	+	beta
9450	Human	O103:H2	*stx1*	+	epsilon1
9574	Human	O111:H8	*stx1+stx2*	+	gamma2/theta
9634	Cow	O111:H8	*stx1+stx2*	-	gamma2/theta
9570	Cow	O111:H8	*stx1+stx2*	+	gamma2/theta
9602	Human	O111:H8	*stx1*	-	gamma2/theta
9455	Unknown	O111:H8	*stx2*	+	gamma2/theta
9340	Human	O103:H25	*stx1*	+	gamma2/theta

FIG 1 Dendrogram of PFGE profiles of 15 O26, O103, and O111 STEC isolates. The similarity of the PFGE profiles was based on the Dice algorithm with 1.5% tolerance. O26:H11 and O111:H11 strains showed a close relationship, grouped in the same cluster, and shared the same *eae* subtype. CVM, Center for Veterinary Medicine.

at the same tRNA locus (*pheU*-tRNA). The O111:H11 strains used in this study also carry β-*eae* as well (Fig. 1). In addition, Konczy et al. (12) and Ziebell et al. (21) demonstrated that STEC O69:H11 was found to be closely related to O26:H11. Additionally, Konczy et al. (12) reported that H25 STEC strains (O103:H25, O119:H25, and O98:H25) and H21 STEC strains (O91:H21, O113:H21, O146:H21, and ONT:H21) were clustered together, separately, based on MLST. These data and our findings provided strong evidence that some STEC strains with common H antigens appear to originate from common ancestors. It is interesting that the four H groups included the serotypes O26:H11, O111:H11, O111:NM, O111:H2, O103:H2, O103:H25, O45:H2, O91:H21, and O113:H21, which have been identified among the most important non-O157 STEC serotypes associated with outbreaks and HUS. Thus, we hypothesize that some clinical and epidemic STEC serotypes with the same H antigens might have evolved from common ancestors, respectively. It is possible that ancestral strains of those H groups share similar or the same genetic background and/or environmental niche that could facilitate acquisition of *stx* and other virulence genes essential to STEC pathogenesis.

Our phylogenetic analyses demonstrated that the H groups were monophyletic while the serogroups were polyphyletic (Fig. 3). Scattered distribution of different O111 strains suggested that strains from individual lineages might have acquired surface an-

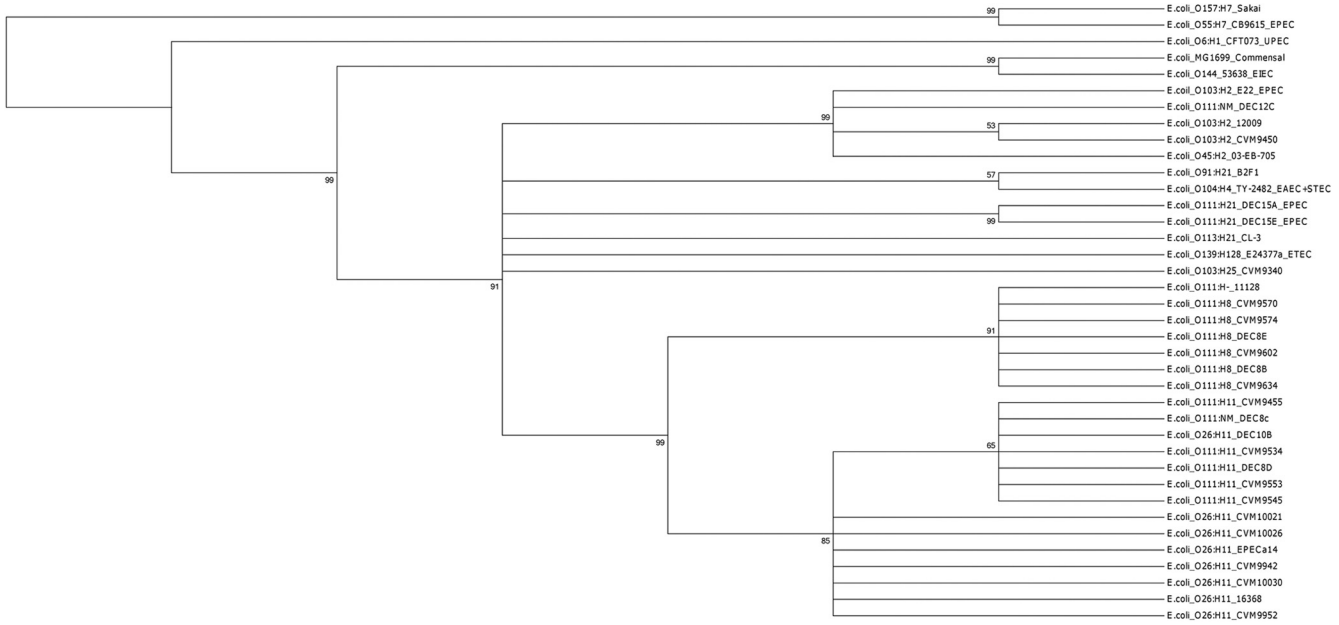


FIG 2 Dendrogram of MLST analyses using *aspC*, *clpX*, *fadD*, *icdA*, *lysP*, *mdh*, and *uidA*. Strains 4865/96 (O145:H28), 101-1 (O−:H10), 5.0959 (O121:H19), DEC12B (O111:H2), and E110019 (O111:H9) were not included in the MLST study because at least one of the selected gene alleles was either absent or only partially present.

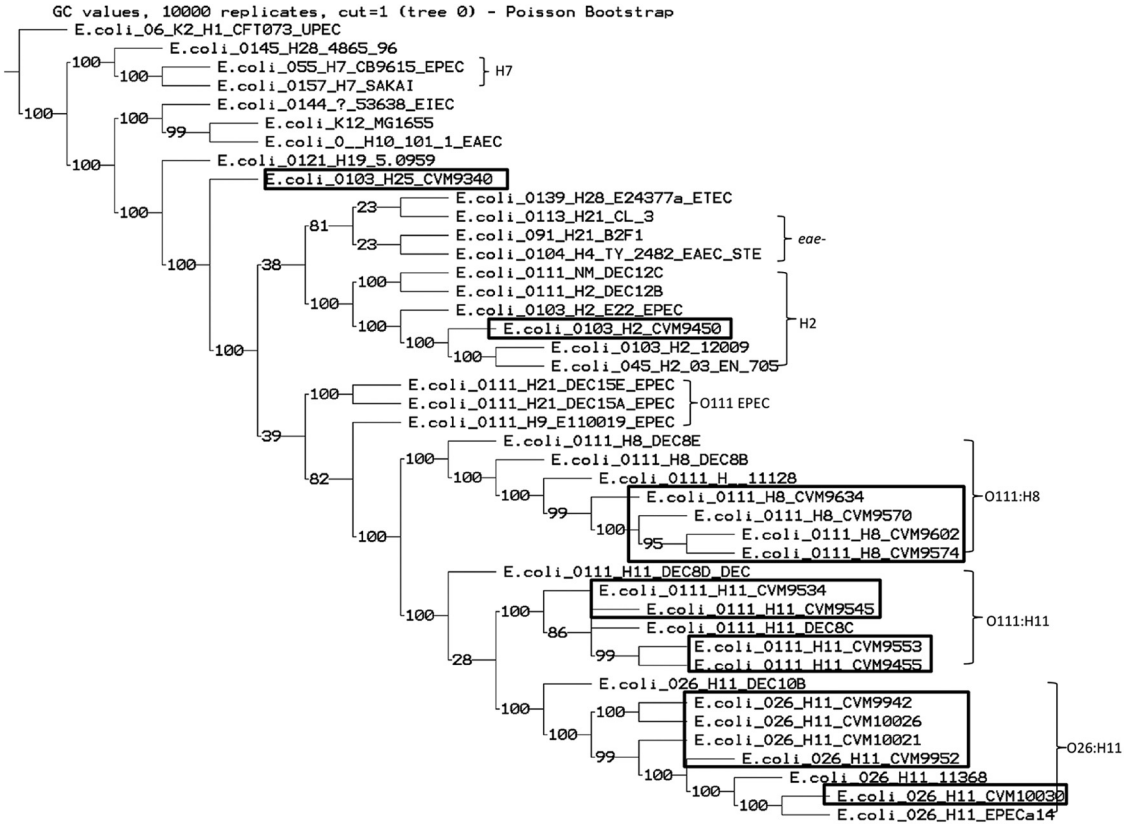


FIG 3 Parsimony phylogenetic tree of 43 *E. coli* strains from diverse pathotypes based on genome-wide single nucleotide polymorphisms (SNPs) with 10,000 iterations. Strains sequenced in this study are shown within boxes. A total of seven subgroups were labeled, as shown in the pairwise distance matrix (Table 2).

tigen genes independently in an ongoing parallel evolutionary process, such as *E. coli* O111:H21 DEC15E (O111 EPEC group), *E. coli* O111:H2 E22 (H2 group), *E. coli* O111:H8 CVM9634 (O111:H8 group), and *E. coli* O111:H11 CVM9545 (O111:H11 group) (Fig. 3). Iguchi et al. (11) suggested that STEC O103:H2, O103:H11, and O103:H25 formed three different lineages by MLST analysis and had distinct *eae* subtypes. The MLST and SNP phylogenetic trees in this study also supported the idea that O103:H2 and O103:H25 were located on different lineages (Fig. 2 and 3). Thus, using just serogroups may cause misleading conclusions about the phylogenetic relatedness and health risks of STEC strains.

Pairwise distance matrix analysis with 2,000 bootstrap iterations (substitution, transitions plus transversions; complete-delete option) (Table 2) was conducted to determine the number of SNP differences (standard deviation) between different selected groups using MEGA 5.05 (17), including H7, *eae* type, H2, O26:H11, and O111 groups. The values of base differences per sequence from averaging overall sequence pairs between groups were shown. The smallest distance value was found between O111:H11 and O26:H11 strains, confirming their close relatedness. Previous studies indicated that O157:H7 evolved from O55:H7 in a series of steps, acquiring the O157 antigen gene cluster and other virulence genes through horizontal gene transfer (HGT) (6, 7, 15, 18). The distance between O157:H7 Sakai and O55:H7 CB9615 was 4,215 SNPs with a standard deviation of 37. Because O26:H11 strains are located in the O111 clade in phylogenetic trees and display a closer relationship with O111:H11 strains than with

TABLE 2 Pairwise distance matrix analysis of six selected groups

Group[a]	No. of SNP differences (SD)					
	H7	*eae* negative	H2	O111 EPEC	O111:H8	O111:H11
H7						
eae negative	57,246 (107)					
H2	58,029 (139)	21,523 (77)				
O111 EPEC	59,530 (105)	23,107 (98)	23,442 (113)			
O111:H8	59,498 (126)	23,993 (75)	22,558 (97)	21,427 (83)		
O111:H11	59,417 (113)	24,157 (98)	22,717 (123)	21,512 (84)	4,324 (35)	
O26:H11	57,176 (108)	21,913 (78)	22,138 (107)	24,045 (102)	6,556 (42)	3,617 (37)

[a] Groups are as shown in Fig. 3.

O111:H8 strains, it is possible that O26:H11 evolved similarly from an ancestral O111:H11 strain by antigenic shift from O111 to O26 (the distance between O26:H11 and O111:H11 groups was 3,617 [Table 2]). Sharing the same niche with other O26 strains may facilitate this genetic exchange. Comparative genomics analysis of O26:H11, O111:H11, and other STEC strains is under way to reveal the possible mechanism.

In conclusion, analyses based on whole-genome-wide SNPs, MLST, and PFGE suggest that on some occasions O serogroups appear not to track evolutionary relatedness among pathogenic *E. coli* strains. Instead, H antigens may be better markers of shared ancestry for some STEC serotypes.

ACKNOWLEDGMENT

The study was supported in part by the Joint Institute for Food Safety and Applied Nutrition (JIFSAN), University of Maryland, College Park, MD.

REFERENCES

1. **Bettelheim KA.** 2007. The non-O157 shiga-toxigenic (verocytotoxigenic) *Escherichia coli*; under-rated pathogens. Crit. Rev. Microbiol. **33**:67–87.
2. **Blanco JE, et al.** 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.
3. **Brooks JT, et al.** 2005. Non-O157 Shiga toxin-producing *Escherichia coli* infections in the United States, 1983–2002. J. Infect. Dis. **192**:1422–1429.
4. **Caprioli A, Morabito S, Brugere H, Oswald E.** 2005. Enterohaemorrhagic *Escherichia coli*: emerging issues on virulence and modes of transmission. Vet. Res. **36**:289–311.
5. **Darling AC, Mau B, Blattner FR, Perna NT.** 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res. **14**:1394–1403.
6. **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.
7. **Feng PC, et al.** 2007. Genetic diversity among clonal lineages within *Escherichia coli* O157:H7 stepwise evolutionary model. Emerg. Infect. Dis. **13**:1701–1706.
8. **Gascuel O.** 1997. BIONJ: an improved version of the NJ algorithm based on a simple model of sequence data. Mol. Biol. Evol. **14**:685–695.
9. **Goloboff P, Nixon FSK.** 2008. TNT, a program for phylogenetic analysis. Cladistics **24**:774–786.
10. **Gouy M, Gascuel GSO.** 2010. SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. Mol. Biol. Evol. **27**:221–224.
11. **Iguchi A, Iyoda S, Ohnishi M.** 2012. Molecular characterization reveals three distinct clonal groups among clinical Shiga toxin-producing *Escherichia coli* strains of serogroup O103. J. Clin. Microbiol. **50**:2894–2900.
12. **Konczy P, et al.** 2008. Genomic O island 122, locus for enterocyte effacement, and the evolution of virulent verocytotoxin-producing *Escherichia coli*. J. Bacteriol. **190**:5832–5840.
13. **Miliwebsky E, et al.** 2007. Prolonged fecal shedding of Shiga toxin-producing *Escherichia coli* among children attending day-care centers in Argentina. Rev. Argent. Microbiol. **39**:90–92.
14. **Ogura Y, et al.** 2009. Comparative genomics reveal the mechanism of the parallel evolution of O157 and non-O157 enterohemorrhagic *Escherichia coli*. Proc. Natl. Acad. Sci. U. S. A. **106**:17939–17944.
15. **Rump LV, et al.** 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.
16. **Scallan E, et al.** 2011. Foodborne illness acquired in the United States—major pathogens. Emerg. Infect. Dis. **17**:7–15.
17. **Tamura K, et al.** 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. **28**:2731–2739.
18. **Tarr PI, et al.** 2000. Acquisition of the rfb-gnd cluster in evolution of *Escherichia coli* O55 and O157. J. Bacteriol. **182**:6183–6191.
19. **Tozzi AE, et al.** 2003. Shiga toxin-producing *Escherichia coli* infections associated with hemolytic uremic syndrome, Italy, 1988–2000. Emerg. Infect. Dis. **9**:106–108.
20. **Tramuta C, Robino P, Oswald E, Nebbia P.** 2008. Identification of intimin alleles in pathogenic *Escherichia coli* by PCR-restriction fragment length polymorphism analysis. Vet. Res. Commun. **32**:1–5.
21. **Ziebell K, et al.** 2008. Applicability of phylogenetic methods for characterizing the public health significance of verocytotoxin-producing *Escherichia coli* strains. Appl. Environ. Microbiol. **74**:1671–1675.
22. **Zwickl DJ.** 2006. Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. Ph.D. thesis. The University of Texas at Austin, Austin, TX.