Applicability of Phylogenetic Methods for Characterizing the Public Health Significance of Verocytotoxin-Producing *Escherichia coli* Strains[▽]

Kim Ziebell,¹ Paulina Konczy,¹ Irene Yong,¹ Shelley Frost,¹ Mariola Mascarenhas,¹
Andrew M. Kropinski,¹ Thomas S. Whittam,² Susan C. Read,¹
and Mohamed A. Karmali^{1*}

Laboratory for Foodborne Zoonoses, Public Health Agency of Canada, Guelph, Ontario N1G 3W4, Canada,¹ and
Microbial Evolution Laboratory, National Food Safety and Toxicology Center, Michigan State University,
East Lansing, Michigan 48824²

Received 16 July 2007/Accepted 20 December 2007

Two phylogenetic methods (multilocus sequence typing [MLST] and a multiplex PCR) were investigated to determine whether phylogenetic classification of verocytotoxin-producing *Escherichia coli* serotypes correlates with their classification into groups (seropathotypes A to E) based on their relative incidence in human disease and on their association with outbreaks and serious complications. MLST was able to separate 96% of seropathotype D and E serotypes from those that cause serious disease (seropathotypes A to C), whereas the multiplex PCR lacked this level of seropathotype discrimination.

The more than 200 serotypes of verocytotoxin-producing *Escherichia coli* (VTEC) (8) that have been isolated from humans (18) may be classified into five seropathotypes (SPTs) based on the relative incidence of the serotypes in human infections and on their association with outbreak infections and serious complications, such as hemolytic-uremic syndrome (7). SPT A comprises serotypes O157:H7 and O157:H–, which commonly cause outbreaks and severe disease, whereas serotypes in SPT E are animal isolates that have never been associated with human disease.

Classification of VTEC serotypes by phylogenetic methods has also demonstrated clustering of some serotypes according to their public health impact (16). Using multilocus sequence typing (MLST), Whittam (16) found that VTEC strains that are associated with outbreaks and/or hemolytic-uremic syndrome correspond to four clonally related clusters referred to as EHEC 1, EHEC 2, STEC 1, and STEC 2 (where EHEC is enterohemorrhagic *E. coli* and STEC is Shiga toxin-producing *E. coli*). Using data from multilocus enzyme electrophoresis, Herzer and colleagues (6) subdivided the ECOR collection (11), comprising a reference group of *E. coli* strains representative of environmental isolates, into phylogenetic groups A, B1, B2, and D. VTEC can be classified into ECOR B1, A, and D phylogroups (PGs) by this approach (2, 3, 5), which, as shown by Clermont and coworkers, may be performed rapidly by a multiplex PCR (1). However, the MLST and the Clermont multiplex PCR methods have not been compared directly.

The objective of this study is to correlate the above-described two phylogenetic methods with the SPT classifica-

tion and to determine whether either phylogenetic method is helpful in predicting the public health significance or SPT of VTEC serotypes isolated from human or nonhuman sources.

The study strains are listed in Table 1 (7). All isolates used in this study were serotyped by using reference O- and H-specific antisera (4) and were tested for *stx* and *eae* (12, 19). The isolates that fell into the EHEC 2 group were further tested for O island 122 (OI-122) (7). The nonmotile O121 and O165 strains were further typed using the flagellar PCR technique (10).

Seven housekeeping genes were amplified from all isolates (*aspC*, *clpX*, *fadD*, *icdD*, *lysP*, *mdhD*, and *uidA*), as previously described (STEC Center; <http://www.shigatox.net/stec/index.html>), by use of their specific primer pairs under standard conditions (13). The amplicons were sequenced using a MegaBACE 500 instrument (GE Health Care, Piscataway, NJ), and the raw sequence data were interpreted with this equipment's software. Data were analyzed further using Lasergene software (Dnastar, Inc., Madison, WI), and sequences were aligned using ClustalX (15). The genes were concatenated in the order listed above, and a phylogenetic tree based on these sequences was constructed with the program MEGA, version 3.1 (9). The trees were calculated with the neighbor-joining algorithm and the Tajima-Nei model, pairwise deletion, gamma 1.0. Bootstrap confidence values were calculated over 1,000 replicate trees. STEC sequence types and allele groups were assigned by the STEC Center (<http://www.shigatox.net/stec/index.html>). A clonal group (CG) designation of 0 indicates an undefined CG.

The ECOR PG was determined by a multiplex PCR-based method as described previously (1).

The clonal diversity of the collection of isolates based on the MLST results is displayed in Fig. 1. Overall, the SPT collection was divisible into 23 different CGs (Table 2). SPT A and B serotypes belonged to five MLST CGs separate from the CGs

* Corresponding author. Mailing address: Laboratory for Foodborne Zoonoses, Public Health Agency of Canada, 110 Stone Rd. W, Guelph, Ontario N1G 3W4, Canada. Phone: (519) 826-2606. Fax: (519) 822-2280. E-mail: Mohamed_Karmali@phac-aspc.gc.ca.

[▽] Published ahead of print on 28 December 2007.

TABLE 1. List of study strains

Strain ^a	Serotype	Source	stx	LEE ^b	MLST CG	MLST ST ^c	ECOR PG	Reference
Strains of serotypes corresponding to SPT A								
9311	O157:H7	Human	1,2	+	11	66	D	7
OK-1	O157:H7	Human	1,2	+	11	66	D	7
EC01-141	O157:H7	Bovine	1,2	+	11	66	D	
279F1	O157:H7	Human	1,2	+	11	66	D	7
278F1	O157:H7	Human	2	+	11	66	D	7
158F2	O157:NM	Human	1,2	+	11	66	D	7
E32511	O157:NM	Human	2	+	11	66	D	7
ER63-94	O157:NM	Human	1,2	+	11	66	D	7
5905	O55:H7	Food	2	+	11	73	D	
Strains of serotypes corresponding to SPT B								
EC94-50	O26:H11	Human	1	+	14	106	B1	
EC95-17	O26:H11	Bovine	1	+	14	106	B1	
EC96-464	O26:H11	Bovine	1	+	14	106	B1	
DEC9F	O26:NM	Human		+	14	106	B1	
DEC8B	O111:H8	Human	1,2	+	14	106	B1	
CL-37	O111:H8	Human	1	+	14	106	B1	
EC93-489	O111:H8	Bovine	1	+	14	106	B1	
202F1(p)	O111:H8	Human	1,2	+	14	106	D	
CL101	O111:NM	Human	1	+	14	106	B1	7
C69F1	O111:NM	Human	1	—	14	106	B1	7
R82F2	O111:NM	Human	1	+	14	106	B1	7
CL106	O121:H19	Human	2	+	24	182	B1	7
274F4	O121:H19	Human	2	+	24	182	B1	7
Z3F1	O121:H19	Human	2	+	24	182	B1	7
N00-6496	O145:NM	Human	1,2	+	12	78	D	7
N01-2051	O145:NM	Human	1	+	12	78	D	7
N02-5149	O145:NM	Human	2	+	12	78	D	7
EC92-185	O103:H2	Human	1	+	17	119	B1	
EC93-204	O103:H2	Bovine	1	+	17	119	B1	
EC97-813	O103:H2	Human	1	+	17	119	B1	
Strains of serotypes corresponding to SPT C								
EC00-850	O91:H21	Bovine	2	—	34	89	B1	
B2F1	O91:H21	Human	2	—	34	89	B1	
EC97-181	O91:H21	Human	2	—	34	89	B1	7
CL3	O113:H21	Human	2	—	30	223	B1	
EC93-474	O113:H21	Bovine	2	—	30	223	B1	
EC97-348	O113:H21	Human	2	—	30	223	B1	
N99-4389	O121:NM	Human	2	+	24	182	B1	7
N99-4390	O121:NM	Human	2	+	24	182	B1	7
N00-4067	O5:NM	Human	1	+	42	688	A	7
N00-4541	O5:NM	Human	1	+	42	687	A	7
N00-4540	O165:H25	Human	2	+	46	253	A	7
Strains of serotypes corresponding to SPT D								
EC93-480	O7:H4	Bovine	2	—	26	85	A	7
EC99-650	O7:H4	Bovine	2	—	26	85	A	
EC92-142	O69:H11	Bovine	1	+	14	104	B1	
EC97-821	O69:H11	Human	1	+	14	104	B1	7
EC92-217	O80:NM	Bovine	2	+	45	245	A	
EC92-248	O80:NM	Bovine	2	+	45	245	A	
EC96-128	O84:H2	Bovine	1	+	20	158	B1	
EC98-521	O84:H2	Bovine	1	+	20	158	B1	
EC92-268	O98:NM	Bovine	2	+	42	246	A	
EC99-345	O98:NM	Bovine	2	+	42	246	A	
N00-9859	O103:H25	Human	1	+	20	159	B1	7
N02-2616	O103:H25	Human	1	+	20	159	B1	7

Continued on following page

TABLE 1—Continued

Strain ^a	Serotype	Source	<i>stx</i>	LEE ^b	MLST CG	MLST ST ^c	ECOR PG	Reference
EC96-371	O113:H4	Bovine	1,2	—	23	171	A	7
EC98-52	O113:H4	Bovine	1,2	—	23	171	A	
N02-0035	O117:H7	Human	1	—	0	689	B1	7
N02-4495	O117:H7	Human	1	—	0	689	B1	7
EC92-267	O119:H25	Human	1	+	20	157	B1	7
EC93-123	O119:H25	Bovine	1	+	20	157	B1	
EC94-374	O119:H25	Bovine	1	+	20	157	B1	
EC92-51	O132:NM	Human	2	—	0	244	D	7
EC92-191	O132:NM	Bovine	2	—	0	244	D	
EC92-374	O132:NM	Bovine	2	—	0	244	D	
EC93-165	O146:H21	Human	1	—	34	89	B1	
EC95-30	O146:H21	Bovine	1	—	34	89	B1	
EC91-62	O165:NM	Bovine	2	+	46	253	A	
EC96-493	O165:NM	Bovine	2	+	46	253	A	
EC92-32	O171:H2	Bovine	2	—	0	130	B1	7
EC02-437	O172:NM	Bovine	2	+	46	252	A	7
A2EV659	O174:H8	Human	1,2	—	19	131	B1	7
Strains of serotypes corresponding to SPT E								
EC96-448	O8:H19	Bovine	1,2	—	43	251	B1	7
EC97-33	O8:H19	Porcine	2E	—	43	254	B1	
EC99-174	O8:H19	Porcine	2E	—	43	254	B1	
EC97-451	O46:H38	Bovine	1,2	—	60	150	B1	7
EC98-46	O46:H38	Bovine	1,2	—	60	150	B1	
EC92-44	O84:NM	Bovine	1	+	20	158	B1	7
EC94-453	O88:H25	Bovine	2	—	60	250	B1	7
EC96-445	O88:H25	Bovine	2	—	60	250	B1	
EC92-413	O98:H25	Bovine	1	+	20	157	B1	
EC93-377	O98:H25	Bovine	1	+	20	157	B1	7
EC93-208	O136:H12	Bovine	1	—	61	248	A	7
EC92-258	O136:NM	Bovine	1	—	61	248	A	7
EC92-104	O153:H31	Bovine	1	—	31	243	B1	7
EC93-569	O153:H31	Bovine	1	—	31	243	B1	
EC92-20	O156:NM	Bovine	2	—	0	242	B1	7
EC92-243	O156:NM	Bovine	2	—	0	242	B1	
EC92-459	O163:NM	Bovine	1,2	—	63	149	B1	7
EC02-238	O177:NM	Bovine	2	+	0	241	A	
Control strains								
K12				—				7
EDL933	O157:H7			1,2	+			7

^a Strains in bold are MLST reference strains.^b A locus of enterocyte effacement (LEE)-positive result is based upon an *eae*-positive result.^c ST, sequence type by MLST.

of SPT D and E, with the exception of one SPT D serotype, O69:H11. This serotype was included in the EHEC 2 (CG 14) group. O69:H11 and other members of the EHEC 2 group were examined for virulence factors (Table 3). Previously, an association between the presence of OI-122 and SPTs linked to epidemic and/or serious disease had been identified (17). The virulence profile for O69:H11 was identical to that of the O26:H11 serotype (Table 3). Additional studies are required to indicate whether this serotype is a threat and could potentially cause serious human illness or whether it lacks yet-unrecognized virulence factors.

The MLST CGs of SPT C serotypes were distinct from the CGs of SPT A and B serotypes, with the exception of CG 24 (Table 2), which contains two serotypes, O121:H19 (SPT B) and O121:NM (SPT C). Using the flagellar (H-antigen-specific) PCR (10), the O121:NM isolate was found to contain the

H19 gene (data not shown) and therefore was not capable of expressing a functional flagellum in vitro. The possible public health significance of this remains to be clarified.

There were three instances of overlap of SPT C and SPT D serotypes within three different CGs (CGs 34, 42, and 46) (Table 2 and Fig. 1). The occurrence of serotypes of different virulence potentials within the same MLST CG suggests that virulence differences are possibly due to horizontal gene transfer of yet-unrecognized virulence factors.

The ECOR phylogrouping shows that the majority (61%) of the serotypes from SPTs B, C, D, and E were PG B1 (Table 1). PGs D and A comprised 12% and 27%, respectively, of the collection of serotypes.

From this study, it is apparent that there is no correlation between the SPT and the ECOR PG in that a specific ECOR PG contains serotypes belonging to several SPT groups.

FIG. 1. Phylogenetic tree of MLST study strains. Letters in parentheses indicate SPT designations. Results were combined when more than one isolate per serotype was of the same sequence type. SC, STEC Center MLST control strains. CG 14 was originally named the EHEC 2 group, CG 17 was originally named the STEC 2 group, CG 34 was originally named the STEC 1 group, CG 30 is split and was originally part of the STEC 1 group, and CG 11 was originally named the EHEC 1 group. Values at left indicate bootstrap confidence values.

Girardeau and coworkers have suggested that the ECOR A PG designation appears to be predictive of low pathogenic potential since there is an absence of the A PG within the SPTs that are associated with disease (A, B, and C) (5). This was confirmed in this study, since there was a paucity of the A PG isolates within SPTs A and B.

In summary, this multiplex PCR method does not provide sufficient discrimination among VTEC serotypes to be useful as a tool to assess their public health significance.

In contrast, using MLST we were able to separate the

majority of the VTEC serotypes associated with severe disease into CGs that are distinct from the CGs of serotypes from SPTs D and E. However, there was some overlap between SPT C and D serotypes within the same MLST CG. This is most likely due to horizontal gene transfer. Thus, while MLST cannot be recommended as a definitive test for predicting virulence, it can contribute to assessing the public health risk of environmental isolates before detailed analyses of virulence factors can be undertaken, especially since it may be possible to perform MLST analyses rapidly

TABLE 2. Correlation of MLST CG and SPT group

MLST CG	No. of serotypes of:				
	SPT A	SPT B	SPT C	SPT D	SPT E
11 (EHEC 1)	2				
12		1			
14 (EHEC 2)		4		1	
17 (STEC 2)		1			
19				1	
20				3	2
23				1	2
24		1	1		
26				1	
30 (STEC 1)			1		
31					1
34 (STEC 1)			1	1	
42			1	1	
43					1
45				1	
46			1	2	
60					2
63					1
X1				1	
X2					1
X3					1
X4				1	
X5				1	

in the future by analyses of single-nucleotide polymorphisms (14).

We thank Teresa Bergholz and Weihong Qi from the STEC Center for training in MLST techniques and the MLST analysis.

The STEC Center is supported in part by the Food and Waterborne Diseases Integrated Research Network (NIH research contract N01-AI-30058).

We also thank Aamir Fazil, Patrick Boerlin, and David Pearl for useful discussions.

TABLE 3. Virulence gene profiles of STEC strains in CG 14

Serotype	No. of isolates	MLST CG	MLST ST ^a	Virulence gene						
				stx ₁	stx ₂	eae	OI-122			
							Z4332	Z4326	Z4321	
O111:H8	2	14	106	+	+	+				
O111:H8	2		106	+	–	+	+	+	+	+
O111:NM	2		106	+	–	+	+	+	+	+
O111:NM	1		106	+	–	–	+	+	+	+
O26:H11	2		106	+	–	+	+	+	+	–
	1		106	+	–	+	+	+	+	–
O26:NM	1		106	–	–	+	+	+	+	–
O69:H11	2		104	+	–	+	+	+	+	–

^a MLST sequence type.

REFERENCES

1. Clermont, O., S. Bonacorsi, and E. Bingen. 2000. Rapid and simple determination of the *Escherichia coli* phylogenetic group. Appl. Environ. Microbiol. **66**:4555–4558.
2. Donnenberg, M. S., and T. S. Whittam. 2001. Pathogenesis and evolution of virulence in enteropathogenic and enterohemorrhagic *Escherichia coli*. J. Clin. Invest. **107**:539–548.
3. Escobar-Paramo, P., O. Clermont, A. B. Blanc-Potard, H. Bui, C. Le Bouguenec, and E. Denamur. 2004. A specific genetic background is required for acquisition and expression of virulence factors in *Escherichia coli*. Mol. Biol. Evol. **21**:1085–1094.
4. Ewing, W. H. 1986. Edward and Ewing's identification of *Enterobacteriaceae*. Elsevier Science Publishing Company, New York, NY.
5. Girardeau, J. P., A. Dalmasso, Y. Bertin, C. Ducrot, S. Bord, V. Livrelli, C. Vernozy-Rozand, and C. Martin. 2005. Association of virulence genotype with phylogenetic background in comparison to different seropathotypes of Shiga toxin-producing *Escherichia coli* isolates. J. Clin. Microbiol. **43**:6098–6107.
6. Herzer, P. J., S. Inouye, M. Inouye, and T. S. Whittam. 1990. Phylogenetic distribution of branched RNA-linked multicopy single-stranded DNA among natural isolates of *Escherichia coli*. J. Bacteriol. **172**:6175–6181.
7. Karmali, M. A., M. Mascarenhas, S. Shen, K. Ziebell, S. Johnson, R. Reid-Smith, J. Isaac-Renton, C. Clark, K. Rahn, and J. B. Kaper. 2003. Association of genomic O island 122 of *Escherichia coli* EDL 933 with verocytotoxin-producing *Escherichia coli* seropathotypes that are linked to epidemic and/or serious disease. J. Clin. Microbiol. **41**:4930–4940.
8. Konowalchuk, J., J. I. Speirs, and S. Stavric. 1977. Vero response to a cytotoxin of *Escherichia coli*. Infect. Immun. **18**:775–779.
9. Kumar, S., K. Tamura, and M. Nei. 2004. MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. Brief. Bioinform. **5**:150–163.
10. Machado, J., F. Grimont, and P. A. Grimont. 2000. Identification of *Escherichia coli* flagellar types by restriction of the amplified *fljC* gene. Res. Microbiol. **151**:535–546.
11. Ochman, H., and R. K. Selander. 1984. Standard reference strains of *Escherichia coli* from natural populations. J. Bacteriol. **157**:690–693.
12. Paton, A. W., and J. C. Paton. 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.
13. Sambrook, J., and D. W. Russell. 2001. Molecular cloning: a laboratory manual, 3rd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
14. Stephens, A. J., F. Huygens, J. Inman-Bamber, E. P. Price, G. R. Nimmo, J. Schooneveldt, W. Munckhof, and P. M. Giffard. 2006. Methicillin-resistant *Staphylococcus aureus* genotyping using a small set of polymorphisms. J. Med. Microbiol. **55**:43–51.
15. Thompson, J. D., D. G. Higgins, and T. J. Gibson. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. **22**:4673–4680.
16. Whittam, T. S. 1998. Evolution of *Escherichia coli* O157:H7 and other Shiga toxin-producing *E. coli* strains, p. 195–209. In J. B. Kaper and A. D. O'Brien (ed.), *Escherichia coli* O157:H7 and other Shiga toxin-producing *E. coli* strains. ASM Press, Washington, DC.
17. Wickham, M. E., C. Lupp, M. Mascarenhas, A. Vazquez, B. K. Coombes, N. F. Brown, B. A. Coburn, W. Deng, J. L. Puente, M. A. Karmali, and B. B. Finlay. 2006. Bacterial genetic determinants of non-O157 STEC outbreaks and hemolytic-uremic syndrome after infection. J. Infect. Dis. **194**:819–827.
18. World Health Organization. 1999. Zoonotic non-O157 Shiga toxin-producing *Escherichia coli* (STEC). Report of a WHO scientific working group meeting. World Health Organization, Geneva, Switzerland.
19. Ziebell, K. A., S. C. Read, R. P. Johnson, and C. L. Gyles. 2002. Evaluation of PCR and PCR-RFLP protocols for identifying Shiga toxins. Res. Microbiol. **153**:289–300.