11-2013

Lineage and Genogroup-Defining Single Nucleotide Polymorphisms of Escherichia coli O157:H7

Woo Kyung Jung
Washington State University College of Veterinary Medicine

James L. Bono
US Meat Animal Research Center, jim.bono@ars.usda.gov

Michael L. Clawson
US Meat Animal Research Center, Mike.Clawson@usda.gov

Shana R. Leopold
Universitätsklinikum Münster

Smriti Shringi
Washington State University College of Veterinary Medicine

See next page for additional authors

AEM
Journals.ASM.org

Lineage and Genogroup-Defining Single Nucleotide Polymorphisms of *Escherichia coli* O157:H7

Woo Kyung Jung,[a] James L. Bono,[b] Michael L. Clawson,[b] Shana R. Leopold,[c] Smriti Shringi,[a] Thomas E. Besser[a]

Department of Veterinary Microbiology and Pathology, Washington State University College of Veterinary Medicine, Pullman, Washington, USA[a]; United States Department of Agriculture, Agricultural Research Service, US Meat Animal Research Center (USMARC), Clay Center, Nebraska, USA[b]; Institut für Hygiene, Universitätsklinikum Münster, Münster, Germany[c]

***Escherichia coli* O157:H7 is a zoonotic human pathogen for which cattle are an important reservoir host. Using both previously published and new sequencing data, a 48-locus single nucleotide polymorphism (SNP)-based typing panel was developed that redundantly identified 11 genogroups that span six of the eight lineages recently described for *E. coli* O157:H7 (J. L. Bono, T. P. Smith, J. E. Keen, G. P. Harhay, T. G. McDaneld, R. E. Mandrell, W. K. Jung, T. E. Besser, P. Gerner-Smidt, M. Bielaszewska, H. Karch, M. L. Clawson, Mol. Biol. Evol. 29:2047–2062, 2012) and additionally defined subgroups within four of those lineages. This assay was applied to 530 isolates from human and bovine sources. The SNP-based lineage groups were concordant with previously identified *E. coli* O157:H7 genotypes identified by other methods and were strongly associated with carriage of specific Stx genes. Two SNP lineages (Ia and Vb) were disproportionately represented among cattle isolates, and three others (IIa, Ib, and IIb) were disproportionately represented among human clinical isolates. This 48-plex SNP assay efficiently and economically identifies biologically relevant lineages within *E. coli* O157:H7.**

Escherichia coli O157:H7 is an important zoonotic pathogen that may cause diarrhea, bloody diarrhea, and hemolytic-uremic syndrome (1, 2). *E. coli* O157:H7 is transmitted to humans indirectly by ingestion of contaminated water or food or directly by contact with infected animals (1, 3). *E. coli* O157:H7 typically produces one or more Shiga toxins (most commonly Stx1, Stx2a, and Stx2c) encoded by lambdoid bacteriophages (4, 5). In U.S. isolates, the Stx1-, Stx2a-, and Stx2c-associated bacteriophages are typically inserted within or adjacent to the conserved chromosomal loci *yehV*, *wrbA* or *argW*, and *sbcB*, respectively (6–8).

Cattle, an important animal reservoir, harbor diverse genotypes of this pathogen as defined by Stx-associated bacteriophage insertion site (SBI) typing. These genotypes include some that are frequently associated with human disease (clinical genotypes) and others that are relatively underrepresented among human clinical isolates (bovine-biased genotypes) (9–11). In addition to these host distribution differences, the SBI genotypes differ in their expression of virulence genes and genes associated with resistance to adverse environmental conditions and in their virulence in animal models (12, 13). Specifically, genes encoding *E. coli* O157:H7 virulence factors, including locus of enterocyte effacement products, enterohemolysin, and pO157 products, showed increased expression in clinical genotypes compared to bovine-biased genotypes. In contrast, genes essential for acid resistance (e.g., *gadA*, *gadB*, and *gadC*) and stress fitness (e.g., *dps*, *osmY*, and *uspA*) were relatively upregulated in bovine-biased genotypes (13). Clinical genotypes of *E. coli* O157:H7 demonstrated more virulence than bovine-biased genotypes in both neonatal piglet and Dutch Belted rabbit models of disease (12). Other molecular subtyping methods that are also strongly influenced by bacteriophage mobility (e.g., pulsed-field gel electrophoresis and octamer-based genome scanning) similarly detect strain types of *E. coli* O157:H7 that are differently represented among isolates from clinical and bovine sources (14, 15).

More recently, single nucleotide polymorphisms (SNPs) identified throughout the *E. coli* O157:H7 genome have been used to characterize the phylogeny of this pathogen (10, 16–19). SNP analysis within the chromosomal backbone is a powerful technique for determining the genetic relatedness and descent of *E. coli* O157:H7 strains in a manner independent of epidemiological data. Initially, large SNP sets are required to inform the phylogeny, but smaller sets of SNP loci can subsequently be used to unequivocally assign isolates to specific clades. The goal of this study was to implement a practical method for determination of SNP alleles and to prospectively evaluate SNP phylogeny for lineages associated with cattle or human disease isolates of *E. coli* O157:H7.

MATERIALS AND METHODS

Bacterial isolates. The strains of *E. coli* O157:H7 used in this study are described in Table S1 in the supplemental material. A total of 342 *E. coli* O157:H7 isolates from cattle and human sources were randomly selected from the strain bank at the Field Disease Investigation Unit at Washington State University, avoiding overrepresentation of individual farms (no more than one isolate per farm per year) or outbreaks (see Table S1). In addition, 67 strains of diverse provenance representing most of the SNP genotypes identified in reference 19 were provided by James L. Bono. Shannon D. Manning (Michigan State University, East Lansing, MI) kindly provided 32 *E. coli* O157:H7 strains representing most of the individual SNP genotypes identified in reference 17, and Nigel P. French (Massey University, Palmerston North, New Zealand) kindly provided 89 isolates from New Zealand cattle and human sources. SBI genotypes of 155 of these strains have been previously reported (20), as indicated in Table S1. Sequences of strain Sakai (GenBank accession no.

Received 1 July 2013 **Accepted** 2 September 2013
Published ahead of print 6 September 2013
Address correspondence to Thomas E. Besser, tbesser@vetmed.wsu.edu.
Supplemental material for this article may be found at http://dx.doi.org/10.1128/AEM.02173-13.

doi:10.1128/AEM.02173-13

NC_002695.1), a fully sequenced Japanese *E. coli* O157:H7 outbreak strain, and *E. coli* O55:H7 strain CB9615 (GenBank accession no. CP001846.1), representing the postulated "ancestral" strain of *E. coli* O157:H7, were used as controls (21, 22). Genomic DNA was extracted using a MagMax total nucleic acid isolation kit (Ambion, Foster City, CA) according to the manufacturer's protocol.

SNP discovery and selection. Clawson et al. (10) described a highly efficient strategy for SNP discovery in *E. coli* O157:H7 by 454 resequencing of DNA pooled from multiple *E. coli* O157:H7 isolates. Each locus where one or more of the resequencing reads differs from the reference sequence represents a candidate polymorphism, and the percentage of the sequencing reads that share the same candidate polymorphism represents the approximate frequency of that polymorphism among the pooled isolates. We previously described five clusters of *E. coli* O157:H7 using the cumulative results of several different genotyping methods: SBI typing (23), LSPA-6 genotyping (24), Q933/Q21 PCRs (25), and an informative *tir* SNP (26). Two of those five clusters were minimally represented in the Clawson SNP discovery effort (10); therefore, we selected a panel of 31 isolates (see Table S2 in the supplemental material) from those clusters and extracted and pooled their genomic DNA for 454 sequencing. The 454 sequencing run produced 306 Mb of sequences for an average read depth of 55, representing on average 1.8× coverage depth of each pooled strain. The 454 reads were assembled against the strain Sakai reference genome (21), with sequence errors corrected as previously reported (16).

Candidate SNPs discovered in this process were screened to remove those identified at loci with poor resequencing coverage (read depth < 10), those that were relatively uncommon among the pooled isolates (alternative allele in <25% of reads at the locus), and those located in O-island rather than backbone loci. After this screening, the remaining 412 SNPs were characterized by genomic location (using strain Sakai as a reference), inter- or intragenic location, and, for intragenic SNPs, synonymous versus nonsynonymous nature. The SNPs included 207 that had been previously published (10, 16, 17, 19) as well as 205 that had not previously been reported. The newly identified SNPs included 58 in intergenic regions and 55 synonymous and 92 nonsynonymous SNPs in annotated open reading frames. SNPs selected for the initial 96-plex assay included 20 designed to identify six of the eight major lineages defined previously (19) and 76 of the novel SNPs identified here. The latter were randomly selected from among the 205 identified, stratified to include approximately equal numbers of SNPs affecting 25% to 49%, 50% to 74%, and 75% to 100% of the read depth.

SNP typing and data analysis. Initially, SNPs were assayed using a 96-plex GoldenGate assay system. Genomic DNA from wild-type *E. coli* O157:H7 isolates and controls was genotyped at the IBEST DNA Sequencing Analysis Core (Department of Biological Sciences, University of Idaho) using the Illumina BeadArray platform. The allele data for the 96 SNP loci were concatenated by physical order along the *E. coli* O157:H7 genome. Subsequently, a subset of SNPs was chosen after elimination of loci used in the 96-plex assay that failed to produce reliable allele reads, removal of loci that provided more than 3-fold redundancy in the identification of some lineage groups, and addition of a key SNP in *tir* (26) to produce the SNP set used in the 48-plex assay (see Table S2 in the supplemental material). The final 48-plex SNP assay included 27 SNPs that are newly published here and 21 SNPs that were reported previously (10, 19). Two different alleles of SNP 544618 were found within one of the lineage groups, so that locus was not considered for lineage determination but is reported here because the locus appears to identify a geographically associated clade.

SBI site genotyping. Stx-encoding bacteriophage insertion (SBI) genotypes were determined by multiplex PCR detecting 12 potential targets (bilateral bacteriophage-chromosomal insertion sites for three phages in four chromosomal sites as well as stx_1, stx_{2a}, and stx_{2c}, as described previously) (20). PCR products were analyzed by capillary electrophoresis (DNA analyzer 3730, LIZ 1200 size standard; Applied Biosystems, Foster City, CA), and data were analyzed with GeneMarker software (SoftGenetics, LLC, State College, PA).

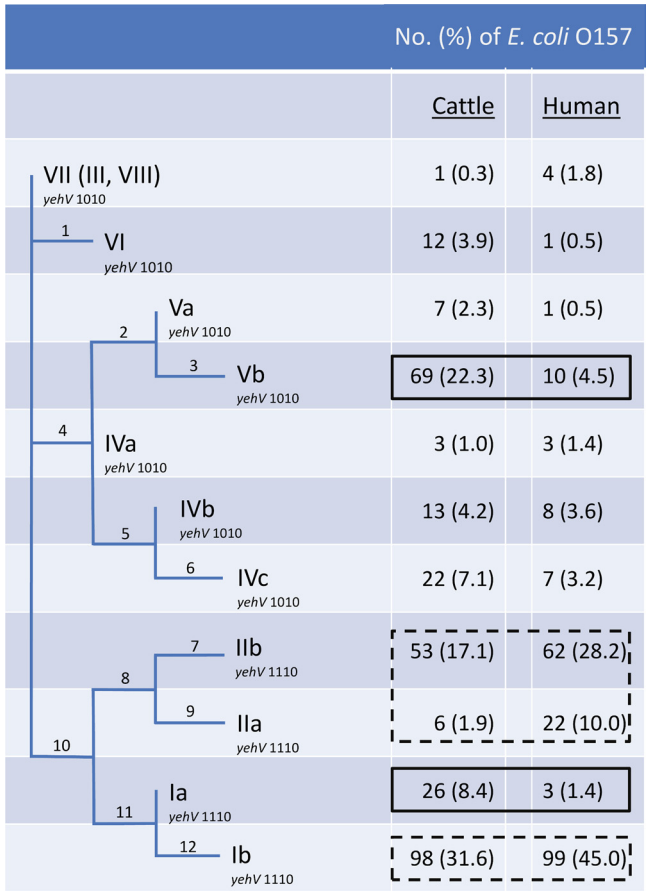

FIG 1 Schematic diagram of SNP genotypes of *E. coli* O157:H7 isolates from cattle and humans to illustrate relative phylogenetic positions (not drawn to scale). Numbers on branches identify the SNP loci that support each bifurcation as shown in Table S2 in the supplemental material. The numbers (percentages) of each of the SNP-defined lineages isolated from cattle and human sources are indicated in the table. Boxed tabulations identify lineages significantly associated with cattle (bovine-biased lineages; solid lines) or human (clinical lineages; dashed lines) sources.

Statistical analysis. The distribution of SNP and SBI genotypes in human and bovine isolates and the associations between the SNP and SBI genotypes were evaluated by χ^2 tests, using SAS software (version 9.2; SAS Institute Inc., Cary, NC).

RESULTS

SNP genotyping. SNPs chosen for inclusion in the 48-plex *E. coli* O157:H7 assay included 25 nonsynonymous SNPs (of which 3 encoded premature stop codons), 18 synonymous SNPs, and 5 intergenic SNPs (see Table S2 in the supplemental material). The assay was applied to 530 *E. coli* O157:H7 human clinical or cattle isolates (see Table S1), and as designed, the concatenated alleles of these isolates segregated into 11 haplotype groups (Fig. 1), corresponding to six of the previously reported eight *E. coli* O157:H7 lineages (I, II, and IV to VII) (19), with additional subdivision of each of lineages I, II, and V into two subgroups designated "a" and "b" and lineage IV into three subgroups designated "a," "b," and "c" (Fig. 1). A total of 112 isolates previously typed by determination of alleles of 175 SNPs were each assigned to the corresponding

TABLE 1 Carriage of Stx-encoding genes by SNP lineages of *E. coli* O157:H7

SNP lineage	No. of *E. coli* O157:H7 isolates carrying indicated Stx-encoding gene(s)							
	stx$_1$, *stx*$_{2a}$	*stx*$_{2a}$	*stx*$_{2c}$	*stx*$_1$, *stx*$_{2c}$	*stx*$_{2a}$, *stx*$_{2c}$	None	*stx*$_1$	*stx*$_1$, *stx*$_{2a}$, *stx*$_{2c}$
Ia		1	12	4	7	4	1	
Ib	159	25		1	1	1	8	3
IIa	2	9			16			
IIb		107	1	1	3	3		
IVa	1			1		2	2	
IVb			18	1	2			
IVc			1	26			1	1
Va			7				1	
Vb			46	26	3	2	1	1
VI			12			1		
VII			1	2		2		
Total	162	142	98	62	32	15	14	5

TABLE 2 Distribution of the most frequently detected SBI genotypes across SNP lineages

SNP lineage	No. (%) of isolates with indicated SBI genotype(s)[a]			
	WY12a	**AY2a, ASY2a**2c, **ASY2a**[b]	**SY2c**, **ASY12c, SY12c**[c]	All others
Ia	0	8 (27.6)	16 (55.2)	5 (17.2)
Ib	155 (78.7)	1 (0.5)	0	41 (20.8)
IIa	0	26 (92.9)	0	2 (7.1)
IIb	0	104 (90.4)	2 (1.7)	9 (7.8)
IVa	0	0	1 (16.7)	5 (83.3)
IVb	0	2 (9.5)	5 (23.8)	14 (66.7)
IVc	0	0	26 (89.7)	3 (10.3)
Va	0	0	5 (62.5)	3 (37.5)
Vb	0	1 (1.3)	72 (91.1)	6 (7.6)
VI	0	0	12 (92.3)	1 (7.7)
VII	0	0	2 (40.0)	3 (60.0)
Total	155	142	141	92

[a] SBI genotypes are named by concatenation of detected Stx-encoding bacteriophage sequences at chromosomal loci *argW* (A), *sbcB* (S), *wrbA* (W), and *yehV* (Y) and specific Stx genes *stx*$_1$ (1), *stx*$_{2a}$ (2a), and *stx*$_{2c}$ (2c). Bold text identifies the conserved bacteriophages and associated Stx genes shared by the SBI genotypes grouped in those columns.

[b] Data include all non-WY12a SBI genotypes previously identified as clinical genotypes (20).

[c] Data include all SBI genotypes previously identified as bovine-biased genotypes (20).

lineage groups by the 48-plex SNP assay reported here. Lineage groups were nonindependently associated with human or bovine hosts of origin ($\chi^2 = 61$, 5 df, $P < 0.001$). Genogroups Vb and Ia were more frequently found among cattle isolates than among clinical isolates ($\chi^2 = 32$, 1 df, $P < 0.001$ and $\chi^2 = 12$, 1 df, $P = 0.001$, respectively), and genogroups Ib, IIa and IIb were overrepresented among clinical isolates compared to cattle isolates ($\chi^2 = 9$, 1 df, $P = 0.002$, $\chi^2 = 18$, 1 df, $P < 0.001$, and $\chi^2 = 10$, 1 df, $P = 0.002$, respectively). The remaining genogroups were relatively sparsely represented in our isolate set; therefore, their potential association with a specific host could not be robustly tested.

SBI site typing. Stx-encoding bacteriophage insertion (SBI) genotypes were determined using multiplex PCR to detect *stx*$_1$, *stx*$_{2a}$, *stx*$_{2c}$, and Stx-associated bacteriophage sequences adjacent to recognized Stx-encoding bacteriophage insertion sites (20). Bacteriophage insertion sites were considered occupied if either the left or the right bacteriophage-bacterial backbone junction or both junctions were amplified, as previously described (20). Genotypes were designated by concatenating the uppercase letters representing occupied bacteriophage insertion sites (A = *argW*, S = *sbcB*, W = *wrbA*, and Y = *yehV*) followed by numbers indicating the Stx variants detected (i.e., 1, 2a, and/or 2c). Isolates in the current study included 35 different SBI genotypes (see Table S1 in the supplemental material) with diverse *stx* contents (Table 1). Cross-tabulation of SNP and the SBI genotypes (Table 2) showed strong association of the most prevalent SBI genotypes with specific SNP lineages. All SBI genotype WY12 isolates were of SNP lineage Ib. A majority (139 of 142, 97.8%) of the other SBI genotypes associated with human disease were also classified in SNP lineages I and II, whereas a majority (123 of 141, 87.2%) of the SBI genotypes overrepresented among cattle isolates were of SNP lineages IV to VI (Table 2).

The Stx1-encoding bacteriophage insertion in *yehV* also demonstrated an interesting distribution across SNP lineages, suggesting a strong correlation with the evolution of *E. coli* O157:H7. Insertion data at this locus can be presented as four concatenated digits of "1" or "0," representing a product or no product, respectively, from PCR amplification of the bacteriophage-*yehV* left junction, the bacteriophage-*yehV* right junction, the bacteriophage-*yehV* alternative right junction, and the intact *yehV* insertion site region (20). Most (91.9%) of the 160 isolates of SNP lineages IV through VII had amplification pattern 1010 and lacked *stx*$_1$ (Fig. 1; see also Table S1 in the supplemental material). In contrast, most (89.2%) SNP genotype I and II isolates had amplification pattern 1110, with sublineages that typically either lacked (95.4% of isolates of lineages Ia, IIa, and IIb) or contained (86.3% of isolates of lineage Ib) *stx*$_1$.

DISCUSSION

We report here a simple and efficient 48-plex SNP typing method that by design robustly classifies *E. coli* O157:H7 isolates into 11 clades corresponding to six previously defined lineages (19), including several subgroups. Lineages III and VIII as defined previously were not targeted in this assay because of the relative scarcity of these lineages within isolates of U.S. origin: Bono et al. (19) identified only a single isolate within lineage III, and lineage VIII represents sorbitol-fermenting human source isolates from Germany that can be distinguished from the non-sorbitol-fermenting clade by simple phenotypic tests. Therefore, this SNP panel was developed so as to group "ancestral" clades including Shiga-toxigenic *E. coli* O55:H7 and *E. coli* O157:H7 lineage VIII together with lineage VII as the root of the tree of SNP lineages defined by this assay (19).

SNP lineages Ia and Vb were markedly overrepresented among isolates obtained from cattle, while lineages Ib, IIa, and IIb were significantly overrepresented in clinical strains. Despite these correlations with source hosts, one or more isolates from all 11 SNP lineages detected by this assay originated from each host species. These results are entirely consistent with the identification of cattle as a reservoir of diverse *E. coli* O157:H7 lineages that result in human disease, albeit at considerably different proportions for different lineages.

Two research groups have previously reported SNP-based

genotyping of *E. coli* O157:H7. Manning et al. (17) used 96 SNP loci to identify 39 genotypes in 9 clades within a panel of >500 clinical *E. coli* O157:H7 isolates. Representative strains of those genotypes and clades were obtained for analysis with the 48-plex SNP assay used here, including 32 isolates uniquely representing one genotype each from clades 1 to 4 and 6 to 9. As expected, there was strong agreement between the two SNP-defined lineage assignments. Isolates of clades 1 to 4 from the work by Manning et al. all typed as lineage Ib, and clades 6, 7, 8, and 9 from the work by Manning et al. typed as lineages Ia, Vb, II, and VII, respectively. SNP lineages IVa to -c, Va, Vb, and VI were not represented in the clades from the work by Manning et al., presumably due to the emphasis on clinical strains (and, as a result, to underrepresentation of bovine-biased strains) in their study. Clades 2 and 3 (lineage Ib) and clade 8 (lineage II) from the work by Manning et al. were the most frequently occurring types in their isolate set, and the corresponding Ib and II lineages were the most frequently occurring among our clinical isolates. Another research group, Bono et al. (19), applied a large (>750-SNP) typing panel to >400 *E. coli* O157 isolates to define 175 genotypes that clustered into eight lineages, I to VIII. Strains ($n = 112$) representing 88 of their 175 SNP genotypes were analyzed using our 48-plex SNP panel, which accurately assigned each isolate to the expected lineage reported by Bono et al., and assigned those in lineages I, II, IV, and V to sublineages. The 48-plex assay utilized here therefore provides grouping information consistent with the Bono SNP-derived phylogeny that could be economically applied wherever the GoldenGate hardware platform or a similar system is available.

It is of considerable research interest to determine the lineages of *E. coli* O157 isolates because of the potentially significant differences in biological behavior they may express, including resistance to adverse environmental conditions, virulence in animal models, and potential for differing levels of virulence in humans (12, 13). These differences may simply result from the lineage-associated differences in the occurrence of *stx* variants. Most (81.6%) isolates of the cattle-associated lineages (Ia and Vb) defined here carry stx_{2c} as their only Stx2-encoding gene. Several studies have suggested that *E. coli* O157:H7 strains that similarly carry stx_{2c} but not stx_{2a} may be less pathogenic than strains that carry stx_{2a} (27–29). For example, strains carrying stx_{2c} are frequently found in isolates from both cattle and healthy people (30, 31) whereas stx_{2a} strains are more frequently isolated from patients with serious symptoms (32). Our results support this association: 81.9% of the clinical isolates in this study tested positive for stx_{2a} versus 51.8% of cattle isolates. In contrast, 42.1% of the cattle isolates in this study carried stx_{2c} but not stx_{2a} compared to only 13.6% of clinical isolates with that Stx-encoding gene pattern. The 51.8% of cattle isolates that carry stx_{2a} alone or together with stx_{2c} therefore apparently represent the strains most likely to result in human infection from this animal reservoir. It would be of considerable interest to compare the severities of clinical disease of humans infected with *E. coli* O157:H7 of different SNP lineages and Stx-encoding gene contents as one method of distinguishing the contribution to virulence of these two factors.

The association of specific stx_2 variant content with certain SNP-defined lineages of *E. coli* O157 suggests that acquisitions and/or excisions of the Stx2 variant-encoding bacteriophages occurred relatively infrequently during the evolution of this pathogen. The most frequently occurring SNP lineages among human clinical isolates exhibited highly consistent SBI for Stx2-encoding phages: 92.9% of lineage Ib isolates had Stx2a-encoding bacteriophage sequences inserted in *wrbA*, and 100% of lineage IIa and 93.9% of lineage IIb isolates had Stx2a-encoding bacteriophage sequences inserted in *argW*. Similarly, bacteriophage sequences inserted in *yehV* (often associated with stx_1) were strongly correlated with the SNP-defined lineages: the predominant clinical lineages Ib, IIa, and IIb as well as the cattle-associated genotype IIa predominantly (330 of 369, 89.4%) share the Stx1-encoding bacteriophage-*yehV* right junction sequence identified by Shaikh and Tarr (7), whereas all the other lineages predominantly (148 of 161, 91.9%) carry the alternative right junction sequence identified by Besser et al. (23). Presumably, the ancestor to lineages I and II as defined by Bono et al. (19), and as identified here by SNP alleles marking branch 10 (Fig. 1; see also Table S2 in the supplemental material), experienced recombination at this bacteriophage-*yehV* junction. The correspondence between polymorphisms in Stx-encoding bacteriophage insertion sites and the lineages defined by chromosomal backbone SNPs emphasizes the surprising stability of at least some of these "mobile elements" in the evolution of *E. coli* O157:H7. The associations between *stx* content and SNP-defined lineages of *E. coli* O157:H7 are clearly also concordant with genotypes derived from other methods used to subtype this pathogen. These include SBI typing (Table 2) as described here but also several other methods that had previously been shown to be concordant with SBI, including LSPA-6, Q allele, and *tir* SNP analyses (9). For example, all isolates of SBI genotype WY12 are of SNP lineage Ib, and 105 of 106 SBI genotype AY2 isolates are of SNP lineage II. On the other hand, the SNP panel includes numerous and diverse minor SBI genotypes within single SNP groups (Table 2) that presumably result from the infrequent bacteriophage insertions or excisions that alter their SBI genotypes but not their lineages as defined by chromosomal backbone SNPs.

As others have previously reported using different genotyping methods (33, 34), SNP lineages of *E. coli* O157:H7 appeared to have pronounced differences in geographical distribution among the seven countries represented; for example, predominant lineages within specific countries of origin included lineage Ia (57% of Australian isolates versus 4.7% elsewhere), lineage Ib (86% of Canadian isolates versus 34.5% elsewhere), lineage IIb (71% of New Zealand isolates versus 11.6% elsewhere), lineage IVb (20.2% of New Zealand isolates versus 0.1% elsewhere), and lineage Vb (51.6% of Japanese isolates versus 12.6% elsewhere) (see Table S1 in the supplemental material). One instance of geographical structuring within a SNP lineage was observed, in that 5 of the 16 Japanese isolates of SNP lineage Vb carried the SNP 544618 allele expected of lineage Va; these were the only isolates in the study with a SNP haplotype intermediate between lineage (sub-)groups. These regional associations suggest that, whether through founder effects or genetic drift or through regional selective pressures, *E. coli* O157:H7 strains have diverged evolutionarily in different parts of the world. The potential for lineage-associated virulence differences accounting, at least in part, for regional differences in the incidence and severity of *E. coli* O157:H7-related disease clearly merits further, more systematic investigation.

In this study, we established a 48-plex SNP typing panel that can economically and rapidly assign *E. coli* O157:H7 isolates to six lineages, including subgroups of four of the lineages. Several of the 11 SNP-defined genogroups were significantly associated with bovine or human animal reservoir and clinical disease sources. The

results were generally concordant with other typing systems, including those based on mobile elements (bacteriophage), representing evidence supporting the important role those elements have played in the evolution of this pathogen. The typing method utilized a platform available at a number of service laboratories and so can be utilized by research laboratories without large investments in specialized hardware to determine SNP genotypes that would cost at least an order of magnitude more for individual determinations by real-time PCR. Further study with more isolates of broader provenance is important for developing a better understanding of the distribution of *E. coli* O157:H7 strains spatially and across diverse hosts, including their association with human diseases.

ACKNOWLEDGMENTS

This project was supported by Agriculture and Food Research Initiative competitive grants no. 2009-04248 and 2010-04487, from the USDA National Institute of Food and Agriculture.

Some of the *E. coli* O157 isolates used in this study were kindly provided by Shannon D. Manning (Michigan State University, East Lansing, MI), Nigel P. French (Institute of Veterinary, Animal, and Biomedical Sciences, Massey University, Palmerston North, New Zealand), and Philip Carter (Institute of Environmental Science and Research Ltd., Poirirua, New Zealand).

REFERENCES

1. **García A, Fox JG, Besser TE.** 2010. Zoonotic enterohemorrhagic *Escherichia coli*: a One Health perspective. ILAR J. **51:**221–232.
2. **Yoon JW, Hovde CJ.** 2008. All blood, no stool: enterohemorrhagic *Escherichia coli* O157:H7 infection. J. Vet. Sci. **9:**219–231.
3. **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.
4. **Nataro JP, Kaper JB.** 1998. Diarrheagenic *Escherichia coli.* Clin. Microbiol. Rev. **11:**142–201.
5. **Melton-Celsa AR, Rogers JE, Schmitt CK, Darnell SC, O'Brien AD.** 1998. Virulence of Shiga toxin-producing *Escherichia coli* (STEC) in orally-infected mice correlates with the type of toxin produced by the infecting strain. Jpn. J. Med. Sci. Biol. **51**(Suppl)**:**S108–S114.
6. **Kotewicz ML, Mammel MK, LeClerc JE, Cebula TA.** 2008. Optical mapping and 454 sequencing of *Escherichia coli* O157: H7 isolates linked to the US 2006 spinach-associated outbreak. Microbiology **154:**3518–3528.
7. **Shaikh N, Tarr PI.** 2003. *Escherichia coli* O157:H7 Shiga toxin-encoding bacteriophages: integrations, excisions, truncations, and evolutionary implications. J. Bacteriol. **185:**3596–3605.
8. **Strauch E, Hammerl JA, Konietzny A, Schneiker-Bekel S, Arnold W, Goesmann A, Puhler A, Beutin L.** 2008. Bacteriophage 2851 is a prototype phage for dissemination of the Shiga toxin variant gene 2c in *Escherichia coli* O157:H7. Infect. Immun. **76:**5466–5477.
9. **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.
10. **Clawson ML, Keen JE, Smith TP, Durso LM, McDaneld 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. doi:10.1186/gb-2009-10-5-r56.
11. **Hancock D, Besser T, Lejeune J, Davis M, Rice D.** 2001. The control of VTEC in the animal reservoir. Int. J. Food Microbiol. **66:**71–78.
12. **Shringi S, Garcia A, Lahmers KK, Potter KA, Muthupalani S, Swennes AG, Hovde CJ, Call DR, Fox JG, Besser TE.** 24 October 2011. Differential virulence of clinical and bovine-biased enterohemorrhagic *E. coli* O157:H7 genotypes in piglet and Dutch Belted rabbit models. Infect. Immun. [Epub ahead of print.] doi:10.1128/IAI.05470-11.
13. **Vanaja SK, Springman AC, Besser TE, Whittam TS, Manning SD.** 2010. Differential expression of virulence and stress fitness genes between *Escherichia coli* O157:H7 strains with clinical or bovine-biased genotypes. Appl. Environ. Microbiol. **76:**60–68.
14. **Böhm H, Karch H.** 1992. DNA fingerprinting of *Escherichia coli* O157:H7 strains by pulsed-field gel electrophoresis. J. Clin. Microbiol. **30:**2169–2172.
15. **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.
16. **Leopold SR, Magrini V, Holt NJ, Shaikh N, Mardis ER, Cagno J, Ogura Y, Iguchi A, Hayashi T, Mellmann A, Karch H, Besser TE, Sawyer SA, Whittam TS, Tarr PI.** 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.
17. **Manning SD, Motiwala AS, Springman AC, Qi W, Lacher DW, Ouellette LM, Mladonicky JM, Somsel P, Rudrik JT, Dietrich SE, Zhang W, Swaminathan B, Alland D, Whittam TS.** 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.
18. **Zhang W, Qi W, Albert TJ, Motiwala AS, Alland D, Hyytia-Trees EK, Ribot EM, Fields PI, Whittam TS, Swaminathan B.** 2006. Probing genomic diversity and evolution of *Escherichia coli* O157 by single nucleotide polymorphisms. Genome Res. **16:**757–767.
19. **Bono JL, Smith TP, Keen JE, Harhay GP, McDaneld TG, Mandrell RE, Jung WK, Besser TE, Gerner-Smidt P, Bielaszewska M, Karch H, Clawson ML.** 2012. Phylogeny of Shiga toxin-producing *Escherichia coli* O157 isolated from cattle and clinically ill humans. Mol. Biol. Evol. **29:**2047–2062.
20. **Shringi S, Ellis C, Katherine K, Brayton KA, Hancock DD, Besser TE.** 11 December 2012. Carriage of stx2a differentiates clinical and bovine-biased strain types of *Escherichia coli* O157. [Epub ahead of print.] PLoS One doi:10.1371/journal.pone.0051572.
21. **Hayashi T, Makino K, Ohnishi M, Kurokawa K, Ishii K, Yokoyama K, Han CG, Ohtsubo E, Nakayama K, Murata T, Tanaka M, Tobe T, Iida T, Takami H, Honda T, Sasakawa C, Ogasawara N, Yasunaga T, Kuhara S, Shiba T, Hattori M, Shinagawa H.** 2001. Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genomic comparison with a laboratory strain K-12. DNA Res. **8:**11–22.
22. **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.
23. **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.
24. **Yang Z, Kovar J, Kim J, Nietfeldt J, Smith DR, Moxley RA, Olson ME, Fey PD, Benson AK.** 2004. Identification of common subpopulations of non-sorbitol-fermenting, beta-glucuronidase-negative *Escherichia coli* O157:H7 from bovine production environments and human clinical samples. Appl. Environ. Microbiol. **70:**6846–6854.
25. **Lejeune JT, Abedon ST, Takemura K, Christie NP, Sreevatsan S.** 2004. Human *Escherichia coli* O157:H7 genetic marker in isolates of bovine origin. Emerg. Infect. Dis. **10:**1482–1485.
26. **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. doi:10.1186/1471-2334-7-98.
27. **Friedrich AW, Bielaszewska M, Zhang WL, Pulz M, Kuczius T, Ammon A, Karch H.** 2002. *Escherichia coli* harboring Shiga toxin 2 gene variants: frequency and association with clinical symptoms. J. Infect. Dis. **185:**74–84.
28. **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.
29. **Nishikawa Y, Zhou Z, Hase A, Ogasawara J, Cheasty T, Haruki K.** 2000. Relationship of genetic type of Shiga toxin to manifestation of bloody diarrhea due to enterohemorrhagic *Escherichia coli* serogroup O157 isolates in Osaka City, Japan. J. Clin. Microbiol. **38:**2440–2442.
30. **Kawano K, Ono H, Iwashita O, Kurogi M, Haga T, Maeda K, Goto Y.** 15 May 2011. 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. [Epub ahead of print.] doi: 10.1007/s10096-011-1283-1.
31. **Nielsen EM, Scheutz F.** 2002. Characterisation of *Escherichia coli* O157 isolates from Danish cattle and human patients by genotyping and presence and variants of virulence genes. Vet. Microbiol. **88:**259–273.
32. **Nakamura H, Ogasawara J, Kita T, Hase A, Nishikawa Y.** 2008. Typing of Stx2 genes of *Escherichia coli* O157 isolates from cattle. Jpn. J. Infect. Dis. **61:**251–252.
33. **Kim J, Nietfeldt J, Ju J, Wise J, Fegan N, Desmarchelier P, Benson AK.** 2001. Ancestral divergence, genome diversification, and phylogeographic variation in subpopulations of sorbitol-negative, beta-glucuronidase-negative enterohemorrhagic *Escherichia coli* O157. J. Bacteriol. **183:**6885–6897.
34. **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.