

rhs Genes Are Potential Markers for Multilocus Sequence Typing of *Escherichia coli* O157:H7 Strains^{∇†}

K. Liu, S. J. Knabel, and E. G. Dudley*

Department of Food Science, The Pennsylvania State University, University Park, Pennsylvania 16802

Received 15 April 2009/Accepted 20 July 2009

DNA sequence-based molecular subtyping methods such as multilocus sequence typing (MLST) are commonly used to generate phylogenetic inferences for monomorphic pathogens. The development of an effective MLST scheme for subtyping *Escherichia coli* O157:H7 has been hindered in the past due to the lack of sequence variation found within analyzed housekeeping and virulence genes. A recent study suggested that *rhs* genes are under strong positive selection pressure, and therefore in this study we analyzed these genes within a diverse collection of *E. coli* O157:H7 strains for sequence variability. Eighteen O157:H7 strains from lineages I and II and 15 O157:H7 strains from eight clades were included. Examination of these *rhs* genes revealed 44 polymorphic loci (PL) and 10 sequence types (STs) among the 18 lineage strains and 280 PL and 12 STs among the 15 clade strains. Phylogenetic analysis using *rhs* genes generally grouped strains according to their known lineage and clade classifications. These findings also suggested that O157:H7 strains from clades 6 and 8 fall into lineage I/II and that strains of clades 1, 2, 3, and 4 fall into lineage I. Additionally, unique markers were found in *rhsA* and *rhsJ* that might be used to define clade 8 and clade 6. Therefore, *rhs* genes may be useful markers for phylogenetic analysis of *E. coli* O157:H7.

Escherichia coli O157:H7 was first described in 1983 as the causative agent of a food-borne outbreak attributed to contaminated ground beef patties (35), and it has subsequently emerged as a very important food-borne pathogen. Diseases caused by *E. coli* O157:H7, such as hemorrhagic colitis and hemolytic uremic syndrome, can be very severe or even life-threatening. Cattle are believed to be the main reservoir for *E. coli* O157:H7 (5, 15, 41), although other animals may also carry this organism (6, 21). Outbreaks are commonly associated with the consumption of beef and fresh produce that come into contact with bovine feces or feces-contaminated environments, such as food contact surfaces, animal hides, or irrigation water (12, 21, 30, 38).

It is well-established that strains of *E. coli* O157:H7 vary in terms of virulence and transmissibility to humans and that strains differing in these characteristics can be distinguished using DNA-based methods (22, 29, 42). For example, octamer-based genome scanning, which is a PCR approach using 8-bp primers, provided the first evidence that there are at least two lineages of O157:H7, termed lineage I and lineage II (22). Strains classified as lineage I are more frequently isolated from humans than are lineage II strains (42). A later refinement of this classification system was coined the lineage-specific polymorphism assay (LSPA), which classified strains based upon the amplicon size obtained using PCRs targeting six chromosomal regions of *E. coli* O157:H7 and assigned a six-digit code based upon the pattern obtained (42). Most strains of lineage I grouped into LSPA type 111111, while the majority of lineage II strains fell into LSPA types 211111, 212111, and 222222.

More recently, it was suggested that LSPA type 211111 strains comprise a separate group called lineage I/II (45).

To gain greater insight into the recent evolution of *E. coli* O157:H7, a method that is more discriminatory than the LSPA method is desirable. Multilocus sequence typing (MLST) is a method that discriminates between strains of a bacterial species by identifying DNA sequence differences in six to eight targeted genes. Satisfactory MLST schemes exist for other bacterial pathogens (28, 43); however, due to the lack of sequence variations in previously targeted gene markers in *E. coli* O157:H7 (13, 33), MLST approaches for subtyping this pathogen have been more difficult to develop. More recently, high-throughput microarray and sequencing platforms have been used to identify hundreds of single nucleotide polymorphisms (SNPs) that are useful for discriminating between strains of *E. coli* O157:H7 during epidemiologic investigations and for drawing phylogenetic inferences (11, 20, 29, 44). Particularly noteworthy, Manning et al. (29) developed a subtyping scheme based upon the interrogation of 32 putative SNP loci. This method separated 528 strains into 39 distinct SNP genotypes, which were grouped into nine statistically supported phylogenetic groups called clade 1 through clade 9. By analyzing the rates of hemolytic uremic syndrome observed in patients infected with strains of clades 2, 7, and 8, it was also concluded that clade 8 strains are more virulent to humans than other strains (29).

One drawback of current DNA sequence-based subtyping schemes for *E. coli* O157:H7 is that they require screening of at least 32 SNP loci. We were interested in asking whether a simpler approach that targets a few informative gene markers could be developed for rapid strain discrimination and phylogenetic determination. A recent analysis of *E. coli* genomes predicted that rearrangement hot spot (*rhs*) genes are under the strongest positive selection of all coding sequences analyzed (34). Therefore, we hypothesized that these genes would

* Corresponding author. Mailing address: 326 Food Science Building, The Pennsylvania State University, University Park, PA 16802. Phone: (814) 867-0439. Fax: (814) 863-6132. E-mail: egd100@psu.edu.

† Supplemental material for this article may be found at <http://aem.asm.org/>.

∇ Published ahead of print on 24 July 2009.

TABLE 1. *E. coli* O157:H7 strains analyzed in the present study

Strain	Classification ^a	Source	Obtained from:	Reference
FR1K1985	Lineage II	Cattle	Andrew Benson	22
FR1K1990	Lineage II	Cattle	Andrew Benson	22
FR1K2000	Lineage II	Cattle	Andrew Benson	22
FR1K2001	Lineage II	Cattle	Andrew Benson	22
FR1K920	Lineage II	Cattle	Andrew Benson	22
FR1K944	Lineage II	Cattle	Andrew Benson	22
NE037	Lineage II	Human	Andrew Benson	22
FDA508	Lineage II	Human	Andrew Benson	22
FDA517	Lineage II	Human	Andrew Benson	22
93-001	Lineage I	Human	Andrew Benson	22
95-003	Lineage I	Human	Andrew Benson	22
FDA507	Lineage I	Human	Andrew Benson	22
FDA518	Lineage I	Human	Andrew Benson	22
NE018	Lineage I	Human	Andrew Benson	22
FR1K523	Lineage I	Human	Andrew Benson	22
FR1K1275	Lineage I	Cattle	Andrew Benson	22
FR1K1986	Lineage I	Cattle	Andrew Benson	22
FR1K1997	Lineage I	Cattle	Andrew Benson	22
Sakai	Clade 1	Human	Wei Zhang	16
TW10022	Clade 1	Human	STEC Center ^b	29
93-111	Clade 2	Human	STEC Center	29
TW11308	Clade 2	Human	STEC Center	29
EDL933	Clade 3	Human	STEC Center	29
TW11346	Clade 3	Human	STEC Center	29
TW11039	Clade 4	Human	STEC Center	29
TW11052	Clade 4	Human	STEC Center	29
TW09109	Clade 6	Human	STEC Center	29
TW11102	Clade 6	Human	STEC Center	29
TW10245	Clade 7	Human	STEC Center	29
TW01663	Clade 7	Human	STEC Center	29
M1-Spinach	Clade 8	Human	Wei Zhang	10
G5101	Clade 9	Human	STEC Center	17
TW07763	Clade 9	Human	STEC Center	29

^a Lineage designations were previously determined by octamer-based genome scanning (22); clade designations were determined by Manning et al. (29).

^b STEC Center, Shiga Toxin *Escherichia coli* Center at Michigan State University.

display significant sequence variations for subtyping O157:H7 strains. The *rhs* genes were first discovered as elements mediating tandem duplication of the *glyS* locus in *E. coli* K-12 (26); however, their function remains unknown. There are nine *rhs* genes within the genome of the prototypical *E. coli* O157:H7 strain Sakai, and these genes are designated *rhsA*, *-C*, *-D*, *-E*, *-F*, *-G*, *-I*, *-J*, and *-K* (see Table S1 in the supplemental material) (16). Three of these nine *rhs* genes, *rhsF*, *-J*, and *-K*, were previously studied by Zhang et al. (44), and a number of SNPs were identified among these genes. However, no studies have been conducted to comprehensively investigate *rhs* genes as markers in an MLST scheme for subtyping *E. coli* O157:H7.

The primary purpose of the present study was to investigate whether there are sufficient DNA sequence variations among *rhs* genes to develop an MLST approach for subtyping *E. coli* O157:H7. In this study, a greater level of DNA sequence variation was observed among *rhs* genes than in gene markers targeted in previous studies (13, 33). Furthermore, phylogenetic analysis using these *rhs* genes generally agreed with the established lineage and clade classifications of O157:H7 strains defined previously. We also wanted to determine whether there is a correlation between the lineage classification of O157:H7 strains (42) and the recently proposed clade classification (29). The present study reports evidence that O157:H7 strains from clade 8 are classified as lineage I/II, which is a different lineage from well-studied *E. coli* O157:H7 outbreak

strains, such as EDL933 and Sakai. Therefore, we suggest that outbreaks of O157:H7 are caused by two lineages of this pathogen, lineage I and lineage I/II.

MATERIALS AND METHODS

Bacterial strains. All *E. coli* O157:H7 strains used in this study are listed in Table 1. Bacterial strains were stored at -80°C in 10% glycerol. When needed, strains were streaked directly onto sorbitol MacConkey agar plates made of Difco MacConkey agar base (Becton, Dickinson and Company, Sparks, MD) supplemented with 1% (wt/vol) D-sorbitol (Alfa Aesar, Ward Hill, MA) and incubated at 37°C overnight.

In silico comparison of O157:H7 *rhs* genes. DNA sequences for *rhs* genes were obtained from whole-genome shotgun (WGS) sequences of 14 *E. coli* O157:H7 strains deposited at the National Center for Biotechnology Information (www.ncbi.nlm.nih.gov). These strains are EC4024 (accession number ABJT00000000), EC4042 (accession number ABHM00000000), EC4045 (accession number ABHL00000000), EC4076 (accession number ABHQ00000000), EC4113 (accession number ABHP00000000), EC4115 (accession number CP001164), EC4196 (accession number ABHO00000000), EC4206 (accession number ABHK00000000), EC4401 (accession number ABHR00000000), EC4486 (accession number ABHS00000000), EC4501 (accession number ABHT00000000), EC508 (accession number ABHW00000000), EC869 (accession number ABHU00000000), and Sakai (accession number BA000007) (16, 23). At the time of manuscript preparation, these genomes were at various stages of assembly and consisted of between one and several hundred contigs. The DNA sequences of the nine *rhs* genes from *E. coli* O157:H7 strain Sakai were extracted using Artemis (version 9.0) (37). The Basic Local Alignment Search Tool (4) was used to identify the corresponding *rhs* gene within each of the whole shotgun sequences. We excluded *rhsG* from the search, as complete sequences of *rhsG* were not found in all WGS sequences at that time. Since *rhs* genes share a high level of sequence identity to one another, each contig encod-

TABLE 2. Sequence variations (SNPs, insertions and deletions) identified by *in silico* comparison of 7 *rhs* genes from 14 *E. coli* O157:H7 strains

Strain	Nucleotide at the indicated gene's position number ^a														Outbreak ^e ST ^d									
	<i>rhsA</i>				<i>rhsC</i>				<i>rhsD</i>		<i>rhsE</i>					<i>rhsF</i>		<i>rhsK</i>						
EC4206	G	C	G	T	C	A	A	G	C	T	C	G	C	T	C	C	C	A	A	A	C	SP	11	
EC4024	SP	12
EC4115	SP	12
EC4113	SP	13
EC4045	SP	13
EC4042	SP	13
EC4196	SP	14
EC4076	SP	15
EC4401	TB	16
EC4486	TJ	16
EC4501	NA	NA
EC508	NA	NA
EC869 ^e	A	.	T	.	T	.	.	G	C	T	17	
Sakai ^b	.	T	.	C	G	G	C	A	18	

^a SNP positions are based upon the corresponding *rhs* gene of the *E. coli* O157:H7 strain Sakai. Vertical numbers represent the nucleotide position of each SNP (i.e., the first column represents nucleotide position 18 of *rhsA*). Identical nucleotides at each SNP locus are indicated by dots. Gaps at each SNP locus are indicated by dashes; #, complete sequences of *rhsC* of EC4501 and EC508 were not found within the GenBank entries; *, there is an insertion from nucleotide positions 1896 to 1949 in *rhsD* of EC508.

^b Sequences of *rhsD* and *rhsE* for strain Sakai were obtained by sequencing in this study. Sequences of the other *rhs* genes for strain Sakai were extracted from GenBank (accession number BA000007) using Artemis (version 9) (37).

^c The first 101 nucleotides of *rhsF* from strain EC869 were missing.

^d I stands for *in silico* comparison; numbers were assigned to different STs.

^e SP stands for the 2006 spinach *E. coli* O157:H7 outbreak (10), TB stands for the 2006 Taco Bell *E. coli* O157:H7 outbreak (9), TJ stands for the 2006 Taco John *E. coli* O157:H7 outbreak (14), Sakai stands for *E. coli* O157:H7 strain Sakai (16), EC508 and EC869 are human fecal and USDA isolates, respectively (20), and are not associated with the aforementioned outbreaks.

ing an *rhs* gene was aligned against the Sakai genome using the Artemis comparison tool (version 6.0) (8) to ensure each gene was given the correct letter designation based upon its location within the genome. The *rhs* genes of the 14 strains were aligned using the Seqman program of the Lasergene software package (DNASTar, Madison, WI).

Amplification of *rhs* genes. Primers for amplifying and sequencing *rhs* genes were designed based upon the *E. coli* O157:H7 strain Sakai genome using Primer 3.0 (<http://frodo.wi.mit.edu/>). PCR templates were prepared by resuspending one colony from sorbitol MacConkey agar plates in 20 µl sterile distilled water. PCR mixtures contained 1.0 to 1.5 µl of this cell suspension as template. The buffer, nucleotides, primer, and *Taq* polymerase (New England Biolabs, Ipswich, MA) concentrations recommended by the manufacturer were used. Primers and cycling conditions are listed in Table S2 of the supplemental material.

DNA sequencing. After PCR, products for sequencing were prepared by adding a 1/20 volume of exonuclease I (10 U/µl; USB Corp., Cleveland, OH) and a 1/20 volume of shrimp alkaline phosphatase (1 U/µl; USB Corp). The mixture was incubated at 37°C for 45 min to degrade the primers and unincorporated deoxynucleoside triphosphates and then at 80°C for 15 min to inactivate the enzymes. Afterwards, PCR products were sent to the Nucleic Acid Facility Center at the Pennsylvania State University (University Park) for sequencing on an ABI Hitachi 3730XL DNA analyzer. In order to obtain the full sequence of each *rhs* gene, sequencing was done with several sequencing primers whose products shared a ca. 150-bp overlap. Sequencing primers for each *rhs* gene are listed in Table S3 of the supplemental material. Sequences of *rhs* genes were aligned and assembled using Seqman. Most of the lengths of the consensus sequences reported are based upon the sequencing of one strand. However, in all cases where SNPs unique to one strain were identified, or where the four color chromatograms did not display unambiguous sequence reads, these regions were sequenced again using a reverse primer to generate double-stranded reads.

LSPA. The LSPA was carried out either in one single multiplex PCR according to the methods of Yang et al. (42) or in two multiplex PCRs according to the methods reported by Ziebell et al. (46) when any of the six genomic markers were not amplified efficiently in one multiplex PCR. Templates were prepared as described above for the amplification of *rhs* genes. When the LSPA was conducted using one single multiplex PCR, a 10-µl aliquot of PCR product was loaded onto a 6% polyacrylamide gel. When the LSPA was conducted using two multiplex PCRs, equal volumes of PCR products from both reaction mixtures were mixed together, and a 10-µl aliquot of this combined mixture was loaded onto a 6% polyacrylamide gel. DNA fragments were visualized after ethidium bromide staining using an EC3 500 BioImaging system (UVP, Upland, CA).

Phylogenetic analysis. SNPs in the eight *rhs* genes of each O157:H7 strain were concatenated and the parsimony-informative (PI) sites (32) were used to construct an unrooted neighbor-joining tree (1,000 replications) using MEGA 4.0 (39). A neighbor-joining tree using allelic profiles of *rhs* genes in O157:H7 was constructed using the Tree Drawing program on the PubMLST website (<http://pubmlst.org/analysis/>).

Nucleotide sequence accession numbers. Nucleotide sequences for *rhs* genes were deposited into GenBank under the accession numbers FJ839695 to FJ839816 and FJ871402 to FJ871406.

RESULTS

In silico comparison of eight *rhs* genes from 14 O157:H7 strains. Previous attempts at developing MLST approaches for *E. coli* O157:H7 (13, 33) were initiated when only two full-genome sequences were available. The recent online release of WGS sequence data for 13 additional strains (EC4206, EC4045, EC4042, EC4115, EC4196, EC4113, EC4076, EC4024, EC4401, EC4486, EC4501, EC508, and EC869) permitted a more comprehensive screen for SNPs. While sequencing and/or assembling errors may exist within these WGS sequences, they still provided a powerful tool for testing hypotheses before sequencing a larger collection of strains. Eight of these 13 strains are from an outbreak traced to contaminated spinach in California in 2006 (10) and 2 are from an outbreak traced to contaminated lettuce from Taco Bell restaurants in 2006 (9), which allowed us to investigate whether *rhs* genes have identical sequences within strains from the same outbreak. We began our study by comparing the sequences of

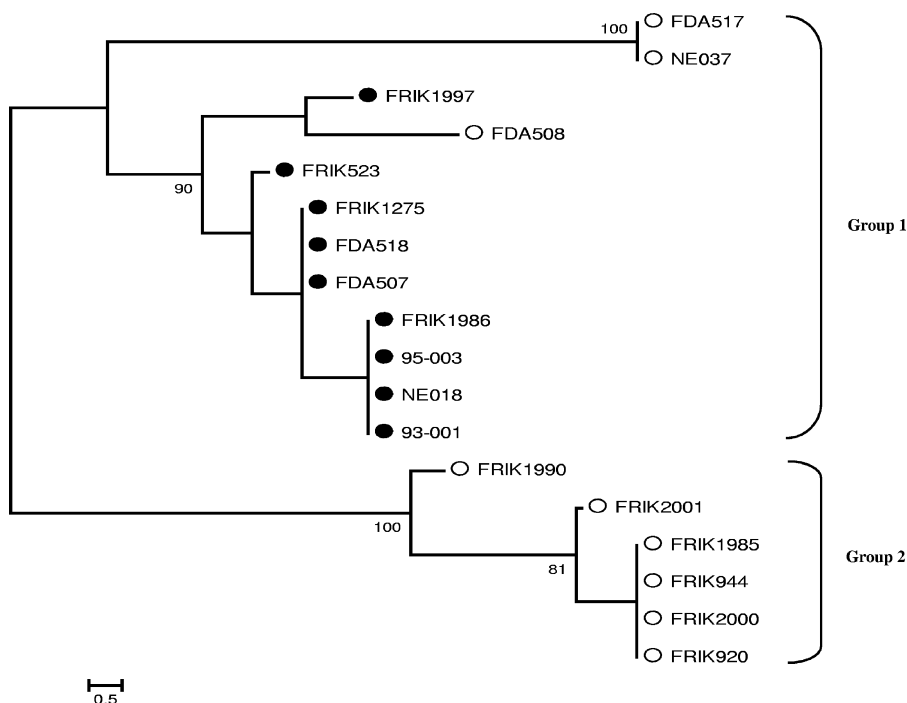


FIG. 1. Unrooted neighbor-joining tree of the 18 lineage strains, based upon the number of differences in parsimony-informative sites among *rhsA*, *-C*, *-D*, *-E*, *-F*, *-J*, and *-K*. SNPs identified in each *rhs* gene (Table 3) were concatenated to represent the SNP sequence type for each O157:H7 strain. Parsimony-informative sites were used to construct the phylogenetic tree. Since sequence information of *rhsI* was not available for all strains, *rhsI* was not included. Bootstrap values (1,000 replications) above 75 are shown at the interior branches. Substitutions per site are indicated by the bar at the bottom. The strains marked by open circles are lineage II strains, and strains marked by solid circles are lineage I strains, whose designations were defined previously by Kim et al. (22) and were confirmed in Table 4.

lineage II strains of cattle origin were LSPA genotypes 222222 (FRIK944, FRIK920, FRIK2000, and FRIK1990), 222221 (FRIK1985), and 222211 (FRIK2001) (Table 4). The three lineage II strains of human origin were classified as LSPA

TABLE 4. LSPA genotypes of *E. coli* O157:H7 strains

Strain	Lineage ^a	LSPA genotype ^b
93-001 ^{†c}	I	111111
95-003 [†]	I	111111
FDA507 [†]	I	111111
FDA518 [†]	I	111111
NE018	I	111111
FRIK523	I	111111
FRIK1275	I	111111
FRIK1986	I	111111
FRIK1997	I	111111
FRIK944	II	222222
FRIK920 [†]	II	222222
FRIK1990 [†]	II	222222
FRIK2000	II	222222
FRIK2001 [†]	II	222211
FRIK1985 [†]	II	222221
NE037 [†]	II	211131
FDA508 [†]	II	211111
FDA517	II	211111

^a Lineage designations were defined by Kim et al. (22).

^b The six gene markers for the LSPA are listed in the following order: *folD-sfmA*, *Z5395*, *yhcG*, *rtcB*, *rbsB*, and *iclR-arp*. Numbers are assigned for each gene marker to represent different alleles.

^c [†], the LSPA genotypes of these strains were determined previously by Yang et al. (42) and were confirmed in the present study.

genotype 211111 (FDA508 and FDA517) and as LSPA genotype 211131 (NE037) (Table 4).

Comparison of seven *rhs* genes from the 15 clade strains.

Recently, Manning et al. (29) used a 32-locus SNP typing scheme to identify nine phylogenetic groups of O157:H7 and designated these groups as clades 1 through 9. As DNA sequence variations were observed in *rhs* genes among lineage strains and phylogenetic analysis using *rhs* genes generally confirmed the phylogeny of the lineage strains determined previously (22), we next tested whether this method would classify strains according to their known clade designation. A strain collection comprised of at least one member of each clade except clade 5 was used. Clade 5 strains were not included, as research since the publication of the paper by Manning et al. (29) suggested that strains originally classified as clade 5 may have been mixed cultures resulting in the distinct phylogeny observed (S. D. Manning, personal communication). Seven *rhs* genes (*rhsA*, *-C*, *-D*, *-E*, *-F*, *-J*, and *-K*) were targeted. We excluded *rhsK*, as only one SNP, found in strain FDA508, was identified previously (Table 3). The full lengths of seven *rhs* genes from the 15 clade strains were obtained, and sequence variations are summarized in Table 5 (see also Tables S4 to S10 in the supplemental material for detailed information). Analysis of the seven *rhs* genes revealed 12 STs among the 15 clade strains, and these STs were designated ST-C1 through ST-C12. Sakai (clade 1), TW10022 (clade 1), and 93-111 (clade 2) shared ST-C1 (Table 5). TW11039 and TW11052 (clade 4) shared ST-C5 (Table 5). Each of the remaining strains pos-

TABLE 5. Summary of allelic types of the seven *rhs* genes from the 15 clade strains

Clade ^a	Strain	Allele type of the indicated gene ^b							ST ^c
		<i>rhsA</i>	<i>rhsC</i>	<i>rhsD</i>	<i>rhsE</i>	<i>rhsF</i>	<i>rhsI</i>	<i>rhsJ</i>	
1	Sakai	1	1	1	1	1	1	1	C1
1	TW10022	1	1	1	1	1	1	1	C1
2	93-111	1	1	1	1	1	1	1	C1
2	TW11308	1	1	2	1	1	1	1	C2
3	EDL933	2	1	1	1	1	1	1	C3
3	TW11346	2	2	3	1	1	1	1	C4
4	TW11039	3	1	1	1	2	2	1	C5
4	TW11052	3	1	1	1	2	2	1	C5
6	TW09109	3	1	4	1	3	3	2	C6
6	TW11102	3	1	1	1	3	3	2	C7
7	TW10245	4	3	1	2	4	2	1	C8
7	TW01663	5	4	1	3	5	NA ^d	1	C9
8	M1-Spinach	6	5	5	4	3	2	1	C10
9	G5101	7	6	6	5	6	2	1	C11
9	TW07763	8	7	7	6	7	2	1	C12

^a Clade designations were defined by Manning et al. (29).

^b Numbers are assigned within each *rhs* gene to represent different allelic types.

^c C stands for clade, and the numbers were assigned to different STs.

^d NA, not available; amplification of *rhsI* from TW01663 was not successful.

sessed a unique ST (Table 5). We identified 280 PL, with *rhsA*, *-C*, *-D*, *-E*, *-F*, *-I*, and *-J* containing 23, 71, 60, 54, 62, 2, and 8 PL, respectively (see Tables S4 to S10 in the supplemental material). Therefore, strains that belong to the same clade may not necessarily share the same *rhs* sequence type. However, an unrooted neighbor-joining tree derived from the data in Table 5 showed that strains of the same clade grouped together as expected (Fig. 2), except for the two strains classified as clade 7.

***E. coli* O157:H7 strains from clade 8 fall into lineage I/II.** We noted from our data that the *rhsA* sequence of strain FDA508, which is characterized as LSPA genotype 211111 (Table 4), is identical to that from the California spinach and the Taco Bell outbreak strains (Tables 2 and 3; see also Table S4 in the supplemental material), which were previously identified as clade 8 (29). We used the SNP-based method of Riordan et al. (36) to confirm in silico that the spinach and

Taco Bell outbreak strains from Table 2 were clade 8, and this method was also used to classify FDA508 as a clade 8 strain (data not shown). Based upon this finding, we hypothesized that clade 8 strains would be LSPA genotype 211111 and would accordingly fall into lineage I/II. The six-gene LSPA (42) showed that the clade 8 strain M1-Spinach from a 2006 outbreak linked to contaminated spinach (10) was LSPA genotype 211111 (Table 6 and Fig. 3). Furthermore, in silico analysis of the six LSPA alleles (*folD-sfmA*, *Z5395*, *yhcG*, *rtcB*, *rbsB*, and *iclR-arp*) revealed that the eight spinach and two Taco Bell outbreak strains shown in Table 2 were also LSPA genotype 211111 (data not shown). Therefore, the clade 8 strains in our study belong to a separate lineage (designated lineage I/II) than other well-studied outbreak strains, such as Sakai, EDL933, and 93-111, which are all LSPA genotype 111111 (Table 6).

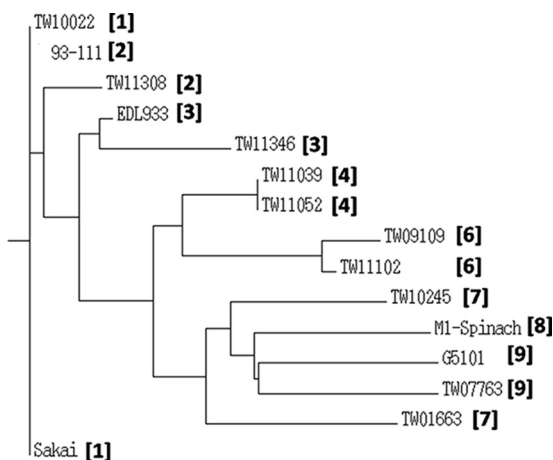


FIG. 2. Unrooted neighbor-joining tree of 15 *E. coli* O157:H7 strains of known clades, based upon the allelic profile of seven *rhs* genes. The tree was constructed using the Tree Drawing program on the PubMLST website (<http://pubmlst.org/analysis>). The clade designation of each strain is indicated by the number in brackets.

TABLE 6. LSPA genotypes of *E. coli* O157:H7 strains

Strain	Clade ^a	LSPA genotype ^b	Lineage ^c
Sakai	1	111111	I
TW10022	1	111111	I
93-111	2	111111	I
TW11308	2	111111	I
EDL933	3	111111	I
TW11346	3	111111	I
TW11039	4	111111	I
TW11052	4	111111	I
TW09109	6	211111	I/II
TW11102	6	211111	I/II
TW10245	7	211111	I/II
TW01663	7	222222	II
M1-Spinach	8	211111	I/II
G5101	9	311111	ND
TW07763	9	212111	II

^a Clade designations were defined by Manning et al. (29).

^b LSPA results, as described in reference 42. The six gene markers are listed in the following order: *folD-sfmA*, *Z5395*, *yhcG*, *rtcB*, *rbsB*, and *iclR-arp*. Numbers were assigned for each gene marker to represent different alleles.

^c Lineage designations were assigned to strains based upon previously defined correlations between LSPA genotype and lineage (42, 45). ND, not defined (the lineage designation for LSPA genotype 311111 has not been defined).

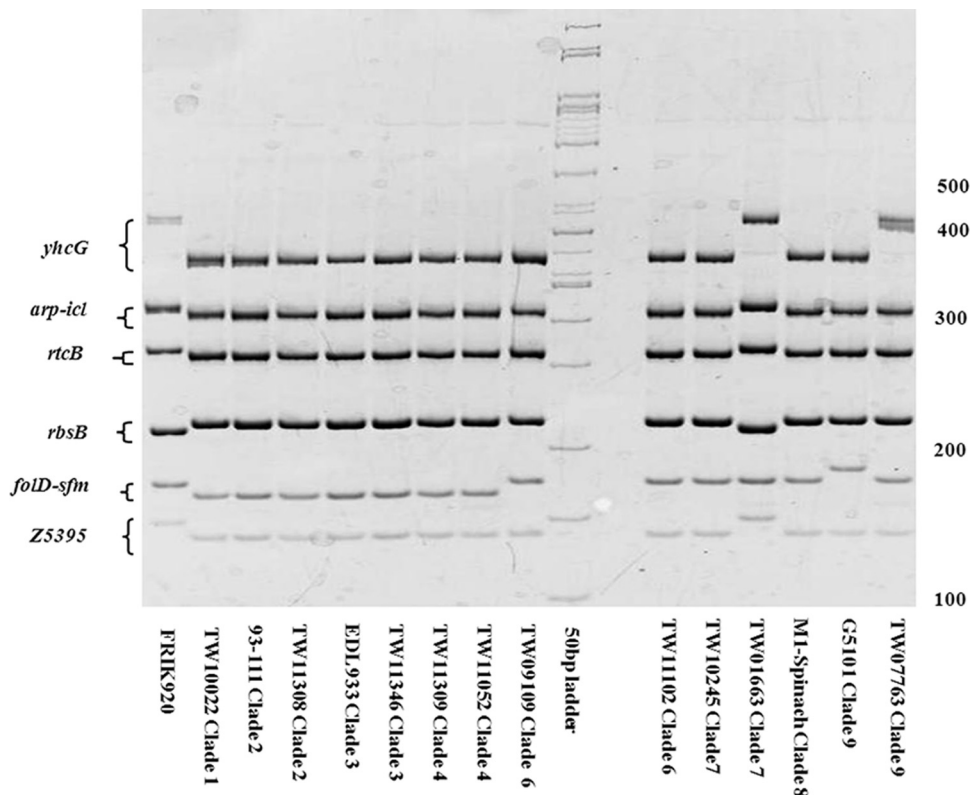


FIG. 3. LSPA of 14 of the clade strains. Fourteen *E. coli* O157:H7 strains from eight clades were subjected to LSPA. The positions of the six gene markers (*folD-sfmA*, *Z5395*, *yhcG*, *rtcB*, *rbsB*, and *iclR-arp*) are indicated on the left. A molecular size marker (50 bp) was loaded in the middle, and sizes are indicated on the right. Strain names and clade designations are indicated below the gel. FRIK920 and EDL933, which were previously reported as LSPA genotype 222222 and LSPA genotype 111111, respectively (45), were used as standards for fragment migration rates. Data from this figure were used to construct Table 6.

Correlation of O157:H7 LSPA genotypes with clade designation. In light of the discovery that clade 8 strains are classified as LSPA genotype 211111 and that no study to date has attempted to investigate the relationship between the LSPA and clade designations, the LSPA genotypes of O157 strains from other clades were determined (Table 6 and Fig. 3). Two major LSPA genotypes were observed, with O157:H7 strains from clades 1, 2, 3, and 4 possessing LSPA genotype 111111 and with strains TW09109 (clade 6), TW11102 (clade 6), and TW10245 (clade 7) typed as LSPA genotype 211111 (Table 6 and Fig. 3). Strains G5101 (clade 9), TW07763 (clade 9), and TW01663 (clade 7) each possessed a unique LSPA genotype and were classified as LSPA genotypes 311111, 212111, and 222222, respectively (Table 6 and Fig. 2). Therefore, LSPA genotype 211111 is not confined to clade 8, but this LSPA group may also include clade 6 strains and possibly a subset of clade 7 strains.

DISCUSSION

Molecular subtyping methods are powerful tools for microbial source tracking and for phylogenetic determination. Several subtyping methods exist for *E. coli* O157:H7 (2, 19, 27), including pulsed-field gel electrophoresis (PFGE) (7), which is generally regarded as the “gold standard” method for subtyping *E. coli* O157:H7 during epidemiologic investigations; however, this method is not suitable for inferring evolutionary relationships between strains (24). DNA se-

quence-based methods, such as MLST, are preferred approaches for inferring phylogeny (1), but previous studies suggested that little sequence variation exists within housekeeping and virulence genes in *E. coli* O157:H7. For example, Noller et al. (33) targeted seven housekeeping genes (*arcA*, *aroE*, *dnaE*, *mdh*, *gnd*, *gapA*, and *pgm*) and two membrane protein genes (*ompA* and *espA*) in their MLST scheme to type a total of 77 O157:H7 isolates. The only sequence variations identified were two SNPs in *ompA*, found in five O157:H7 isolates. In another study, Foley et al. (13) chose one housekeeping gene (*uidA*) and three virulence genes as their targets (*eaecA*, *hlyA*, and *fliC*), and they examined 92 O157:H7 isolates from various sources. They found only five MLST types among the 92 O157:H7 isolates, compared with 72 distinct profiles by PFGE. The remarkable lack of sequence variation observed by these researchers could be explained by two theories that are not mutually exclusive. First, the beta-glucuronidase-negative, sorbitol-negative *E. coli* O157:H7 strains are believed to have emerged from a common ancestor perhaps as early as 7,000 years ago (25), which may not have provided sufficient time for SNPs in housekeeping and virulence genes to accumulate to levels necessary for MLST to be effective. Second, gene markers used in these previous studies may not be under strong enough positive selective pressure, and therefore allelic variants are not selected for at an appreciable rate for these gene markers to be useful in revealing the short-term evolution history of *E. coli* O157:H7. While MLST schemes are also used

for pathogen trace-back studies during outbreaks, our method clearly does not have the ability to discriminate between all unrelated strains and thus is not superior to PFGE for this purpose.

We examined several *rhs* genes which were identified as the genes that are under the strongest positive selection in a survey of *E. coli* genomes (34). Our study showed that these *rhs* genes possessed significantly more sequence variations than house-keeping and virulence genes used in previous studies (13, 33) (Tables 2, 3, and 5). While we appreciate that genes under positive selection are generally avoided for the reconstruction of long-term evolutionary events, we suggest that *rhs* genes may still be useful for revealing short-term evolution history. In the present study, phylogenetic analysis results for the 18 lineage O157:H7 strains using *rhs* genes generally agreed with the phylogeny determined previously (22), although the three O157:H7 strains of human origin from lineage II (NE037, FDA508, and FDA517) clustered together with the nine lineage I strains (Fig. 1). However, using the six-gene LSPA we determined that two of these strains are in the recently described lineage I/II and that NE037 has an LSPA genotype (211131) distinct from lineage I, I/II, and II. Our MLST scheme correctly clustered together strains with LSPA genotype 111111, which is the most prevalent LSPA genotype in lineage I, and grouped together LSPA genotypes 222222, 222221, and 222211, which belong to lineage II (45). Our data suggest that NE037 (LSPA genotype 211131) falls into lineage I/II; however, further studies are needed to confirm this. Additionally, a phylogenetic tree (Fig. 2) constructed from the *rhs* allelic profiles (Table 5) mainly classified the strains screened according to their known clade structure, again supporting the argument that this method accurately predicts short-term evolution of *E. coli* O157:H7. While different STs were identified between strains from the same clade, this is consistent with the same observation of Manning et al. (29).

Two observations with the clade strains studied (Table 5 and Fig. 2) appear to contradict our argument that *rhs* allelic types can be used to infer short-term evolution of *E. coli* O157:H7; however, we do not believe these to be problematic. First, we found that the two clade 7 strains we studied were placed in distinct locations within the phylogenetic tree (Fig. 2). This observation was not unexpected, as the LSPA revealed that these two strains classified as two different LSPA genotypes (211111 and 222222). Additionally, in silico analysis of the WGS sequence of strain EC869 (Table 2) indicates that while it classifies as a clade 7 strain (M. K. Mammel, personal communication), it is LSPA genotype 222223 (K. Liu, unpublished data). Therefore, further study is needed to investigate whether there are indeed multiple LSPA genotypes contained within strains currently defined as clade 7, and if so this might suggest that additional SNPs should be incorporated into current methods to separate these strains from one another. Second, the two clade 1 strains, Sakai and TW10022, shared the same sequence type with 93-111, a clade 2 strain (Table 5). Recently, Riordan et al. (36) targeted SNPs in four genes and developed a quick subtyping method to discriminate clade 1 from clade 2 strains. These observations suggest clades 1 and 2 are highly related, and better methods are needed to differentiate these two clades.

At this point in time, due to the limited number of strains screened, we do not know whether our method identifies a greater number of STs than the SNP-based approaches previously reported (11, 29). Therefore, we plan to sequence *rhs* genes from a larger number of O157:H7 strains, compare the results to these other methods, and use high-throughput sequencing of *E. coli* O157:H7 genomes to identify additional targets for phylogenetic analysis.

In addition to the above observations, the present study also provided data suggesting that O157:H7 strains from clades 6, 7, and 8 are members of lineage I/II. The original classification of the LSPA genotype 211111 grouped these strains into lineage II (42), a group more commonly associated with strains that are of low virulence to humans. Through full-genome phylogenetic analysis, Zhang et al. (45) defined an additional group, lineage I/II, and suggested that this lineage comprises LSPA genotype 211111. Their data also suggest that lineage I/II (LSPA genotype 211111) strains are more closely related to lineage I strains (LSPA genotype 111111) than they are to lineage II strains (LSPA genotypes 222222, 222211, and 222212) (45). Our data support this view and suggest that like lineage I strains, lineage I/II strains are also responsible for causing outbreaks. As only a limited number of strains were analyzed in the present study, we are now screening a larger collection of O157:H7 strains to determine the prevalence of the lineage I/II genotype among sporadic and outbreak strains. Comparative genomic analysis of lineage I, I/II, and II strains may lend insights into differences in pathogenicity and host specificity of these strains.

Lastly, this study identified genetic markers that may be useful for differentiating strains of different lineages and clades. First, as noted previously by Zhang et al. (45), we showed that the complete gene for *rhsI* is not present in lineage II strains, providing further evidence that this marker might be useful for separating these strains from lineages I and I/II. Furthermore, the present study also identified possible SNP markers for identifying clade 6 and clade 8. Four SNPs identified in *rhsA* at nucleotide positions 3468, 3478, 3479, and 3516 (see Table S1 in the supplemental material) could potentially be used as markers for clade 8. These four SNPs were observed in the spinach and Taco Bell outbreak strains (Table 5) as well as in strain FDA508 (Table 6), which were all confirmed to be clade 8 by either sequencing or in silico comparison of ECs2357 (36) (data not shown). Eight unique SNPs in *rhsJ* were identified for the two clade 6 strains, TW09109 and TW11102 (see Table S7 in the supplemental material). Interestingly, NE037 and FDA517 also possess these eight SNPs in *rhsJ* (Table 6). The 32 SNPs defined by Manning et al. (29) need to be screened in order to determine whether NE037 and FDA517 also belong to clade 6.

It is currently unclear why *rhs* genes are under strong positive selection. The function(s) of *rhs* genes is still unknown; however, several studies have shed light on this puzzle. The mature proteins are predicted to be extracellular (18), and phenotypes reported include promotion of the intestinal colonization of *E. coli* in calves (40), the biogenesis of group 2 capsules (31), and increasing the resistance of *E. coli* to the biocide polyhexamethylene biguanide (3). Taking all these together, we hypothesize that Rhs proteins promote survival of *E. coli* during intestinal transit in some undefined manner and that their extracellular nature results in a strong positive se-

lective pressure from the host's immune system. Rhs proteins may also modulate host specificity of *E. coli* O157:H7 and other members of *Escherichia*; however, this idea is completely speculative at this point in time.

In conclusion, our study identified DNA sequence variations in seven *rhs* genes (*rhsA*, *-C*, *-D*, *-E*, *-F*, *-I*, and *-J*) that could potentially serve as markers for subtyping *E. coli* O157:H7. Future studies are required to test this MLST scheme on a larger collection of diverse *E. coli* O157:H7 strains of known phylogenies. Additionally, the present study also suggested that O157:H7 strains from clade 8 fall into lineage I/II and that unique markers in *rhsA* and *-J* could be used for distinguishing clade 8 and clade 6. A larger collection of *E. coli* O157:H7 strains should be tested to determine whether these SNPs are unique to clade 8 and clade 6.

ACKNOWLEDGMENTS

This work was supported by United States Department of Agriculture Cooperative State Research, Education, and Extension Service grant 2008-34163-19283 (E.G.D. and S.J.K.) and by start-up funds from the Department of Food Science and the College of Agricultural Sciences at the Pennsylvania State University (E.G.D.).

We thank Andrew Benson at the University of Nebraska—Lincoln, Thomas Whittam and Beth Whittam at Michigan State University, and Wei Zhang at the Illinois Institute of Technology for kindly providing *E. coli* O157:H7 strains. We thank Thomas Cebula (Johns Hopkins University) for information on the sequenced strains shown in Table 2 and Mark K. Mammel (U.S. FDA Center for Food Safety and Applied Nutrition) for the observation that strain EC869 is classified as clade 7. We also thank Bindhu Verghese for helping with the phylogenetic analysis of *rhs* sequences and Robert Roberts, Jia Wen, Amrita Puri, and Chun Chen, all at the Pennsylvania State University, for reading the manuscript prior to submission.

REFERENCES

- Achtman, M. 2008. Evolution, population structure, and phylogeography of genetically monomorphic bacterial pathogens. *Annu. Rev. Microbiol.* **62**:53–70.
- Ahmed, R., C. Bopp, A. Borczyk, and S. Kasatiya. 1987. Phage-typing scheme for *Escherichia coli* O157:H7. *J. Infect. Dis.* **155**:806–809.
- Allen, M. J., G. F. White, and A. P. Morby. 2006. The response of *Escherichia coli* to exposure to the biocide polyhexamethylene biguanide. *Microbiology* **152**:989–1000.
- Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**:3389–3402.
- Armstrong, G. L., J. Hollingsworth, and J. G. Morris, Jr. 1996. Emerging foodborne pathogens: *Escherichia coli* O157:H7 as a model of entry of a new pathogen into the food supply of the developed world. *Epidemiol. Rev.* **18**:29–51.
- Beutin, L., D. Geier, H. Steinruck, S. Zimmermann, and F. Schütz. 1993. Prevalence and some properties of verotoxin (Shiga-like toxin)-producing *Escherichia coli* in seven different species of healthy domestic animals. *J. Clin. Microbiol.* **31**:2483–2488.
- Byrne, C. M., I. Erol, J. E. Call, C. W. Kaspar, D. R. Buege, C. J. Hiemke, P. J. Fedorka-Cray, A. K. Benson, F. M. Wallace, and J. B. Luchansky. 2003. Characterization of *Escherichia coli* O157:H7 from downer and healthy dairy cattle in the upper midwest region of the United States. *Appl. Environ. Microbiol.* **69**:4683–4688.
- Carver, T. J., K. M. Rutherford, M. Berriman, M. A. Rajandream, B. G. Barrell, and J. Parkhill. 2005. ACT: the Artemis Comparison Tool. *Bioinformatics* **21**:3422–3423.
- Centers for Disease Control and Prevention. 2006. Multistate outbreak of *E. coli* O157 infections, November–December 2006. Centers for Disease Control and Prevention, Atlanta, GA. <http://www.cdc.gov/ecoli/2006/december/121406.htm>.
- Centers for Disease Control and Prevention. 2006. Ongoing multistate outbreak of *Escherichia coli* serotype O157:H7 infections associated with consumption of fresh spinach—United States. *MMWR Morb. Mortal. Wkly. Rep.* **55**:1045–1046.
- Clawson, M., J. Keen, T. Smith, L. Durso, T. McDaniel, R. Mandrell, M. Davis, and J. Bono. 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.
- Elder, R. O., J. E. Keen, G. R. Siragusa, G. A. Barkoocy-Gallagher, M. Koochmarai, and W. W. Laegreid. 2000. Correlation of enterohemorrhagic *Escherichia coli* O157 prevalence in feces, hides, and carcasses of beef cattle during processing. *Proc. Natl. Acad. Sci. USA* **97**:2999–3003.
- Foley, S. L., S. Simjee, J. Meng, D. G. White, P. F. McDermott, and S. Zhao. 2004. Evaluation of molecular typing methods for *Escherichia coli* O157:H7 isolates from cattle, food, and humans. *J. Food Prot.* **67**:651–657.
- Food and Drug Administration. 2007. FDA and states closer to identifying source of *E. coli* contamination associated with illnesses at Taco John's restaurants. <http://www.fda.gov/bbs/topics/NEWS/2007/NEW01546.html>.
- Hancock, D. D., T. E. Besser, M. L. Kinsel, P. I. Tarr, D. H. Rice, and M. G. Paros. 1994. The prevalence of *Escherichia coli* O157:H7 in dairy and beef cattle in Washington State. *Epidemiol. Infect.* **113**:199–207.
- Hayashi, T., K. Makino, M. Ohnishi, K. Kurokawa, K. Ishii, K. Yokoyama, C.-G. Han, E. Ohtsubo, K. Nakayama, T. Murata, M. Tanaka, T. Tobe, T. Iida, H. Takami, T. Honda, C. Sasakawa, N. Ogasawara, T. Yasunaga, S. Kuhara, T. Shiba, M. Hattori, and H. Shinagawa. 2001. Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genomic comparison with a laboratory strain K-12. *DNA Res.* **8**:11–22.
- Hayes, P. S., K. Blom, P. Feng, J. Lewis, N. A. Strockbine, and B. Swaminathan. 1995. Isolation and characterization of a β -D-glucuronidase-producing strain of *Escherichia coli* serotype O157:H7 in the United States. *J. Clin. Microbiol.* **33**:3347–3348.
- Hill, C. W., C. H. Sandt, and D. A. Vlazny. 1994. Rhs elements of *Escherichia coli*: a family of genetic composites each encoding a large mosaic protein. *Mol. Microbiol.* **12**:865–871.
- Iyoda, S., A. Wada, J. Weller, S. J. Flood, E. Schreiber, B. Tucker, and H. Watanabe. 1999. Evaluation of AFLP, a high-resolution DNA fingerprinting method, as a tool for molecular subtyping of enterohemorrhagic *Escherichia coli* O157:H7 isolates. *Microbiol. Immunol.* **43**:803–806.
- Jackson, S. A., M. K. Mammel, I. R. Patel, T. Mays, T. J. Albert, J. E. LeClerc, and T. A. Cebula. 2007. Interrogating genomic diversity of *E. coli* O157:H7 using DNA tiling arrays. *Forensic Sci. Int.* **168**:183–199.
- Jay, M. T., M. Cooley, D. Carychao, G. W. Wiscomb, R. A. Sweitzer, L. Crawford-Miksza, J. A. Farrar, D. K. Lau, J. O'Connell, A. Millington, R. V. Asmundson, E. R. Atwill, and R. E. Mandrell. 2007. *Escherichia coli* O157:H7 in feral swine near spinach fields and cattle, central California coast. *Emerg. Infect. Dis.* **13**:1908–1911.
- Kim, J., J. Niefeldt, and A. K. Benson. 1999. Octamer-based genome scanning distinguishes a unique subpopulation of *Escherichia coli* O157:H7 strains in cattle. *Proc. Natl. Acad. Sci. USA* **96**:13288–13293.
- Kotewicz, M. L., M. K. Mammel, J. E. LeClerc, and T. A. Cebula. 2008. Optical mapping and 454 sequencing of *Escherichia coli* O157:H7 isolates linked to the US 2006 spinach-associated outbreak. *Microbiology* **154**:3518–3528.
- Laing, C., C. Pegg, D. Yawney, K. Ziebell, M. Steele, R. Johnson, J. E. Thomas, E. N. Taboada, Y. Zhang, and V. P. Gannon. 2008. Rapid determination of *Escherichia coli* O157:H7 lineage types and molecular subtypes by using comparative genomic fingerprinting. *Appl. Environ. Microbiol.* **74**:6606–6615.
- Leopold, S. R., V. Magrini, N. J. Holt, N. Shaikh, E. R. Mardis, J. Cagno, Y. Ogura, A. Iguchi, T. Hayashi, A. Mellmann, H. Karch, T. E. Besser, S. A. Sawyer, T. S. Whittam, and P. I. Tarr. 2009. A precise reconstruction of the emergence and constrained radiations of *Escherichia coli* O157 portrayed by backbone concatenomic analysis. *Proc. Natl. Acad. Sci. USA* **106**:8713–8718.
- Lin, R. J., M. Capage, and C. W. Hill. 1984. A repetitive DNA sequence, rhs, responsible for duplications within the *Escherichia coli* K-12 chromosome. *J. Mol. Biol.* **177**:1–18.
- Lindstedt, B. A., E. Heir, E. Gjernes, T. Vardund, and G. Kapperud. 2003. DNA fingerprinting of Shiga-toxin producing *Escherichia coli* O157 based on multiple-locus variable-number tandem-repeats analysis (MLVA). *Ann. Clin. Microbiol. Antimicrob.* **2**:12.
- Maiden, M. C., J. A. Bygraves, E. Feil, G. Morelli, J. E. Russell, R. Urwin, Q. Zhang, J. Zhou, K. Zurth, D. A. Caugant, I. M. Feavers, M. Achtman, and B. G. Spratt. 1998. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc. Natl. Acad. Sci. USA* **95**:3140–3145.
- Manning, S. D., A. S. Motiwala, A. C. Springman, W. Qi, D. W. Lacher, L. M. Ouellette, J. M. Mladonicky, P. Somsel, J. T. Rudrik, S. E. Dietrich, W. Zhang, B. Swaminathan, D. Alland, and T. S. Whittam. 2008. Variation in virulence among clades of *Escherichia coli* O157:H7 associated with disease outbreaks. *Proc. Natl. Acad. Sci. USA* **105**:4868–4873.
- Marina, S., and O. Joseph. 2004. Irrigation water as source of foodborne pathogens on fruit and vegetables. *J. Food Prot.* **67**:2839–2849.
- McNulty, C., J. Thompson, B. Barrett, L. Lord, C. Andersen, and I. S. Roberts. 2006. The cell surface expression of group 2 capsular polysaccharides in *Escherichia coli*: the role of KpsD, RhsA and a multi-protein complex at the pole of the cell. *Mol. Microbiol.* **59**:907–922.
- Nei, M., and S. Kumar. 2000. Phylogenetic inference: maximum parsimony methods, p. 119. In M. Nei and S. Kumar (ed.), *Molecular evolution and phylogenetics*. Oxford University Press, New York, NY.
- Noller, A. C., M. C. McEllistrem, O. C. Stine, J. G. Morris, Jr., D. J. Boxrud,

- B. Dixon, and L. H. Harrison. 2003. Multilocus sequence typing reveals a lack of diversity among *Escherichia coli* O157:H7 isolates that are distinct by pulsed-field gel electrophoresis. *J. Clin. Microbiol.* **41**:675–679.
34. Petersen, L., J. P. Bollback, M. Dimmic, M. Hubisz, and R. Nielsen. 2007. Genes under positive selection in *Escherichia coli*. *Genome Res.* **17**:1336–1343.
35. Riley, L. W., R. S. Remis, S. D. Helgerson, H. B. McGee, J. G. Wells, B. R. Davis, R. J. Hebert, E. S. Olcott, L. M. Johnson, N. T. Hargrett, P. A. Blake, and M. L. Cohen. 1983. Hemorrhagic colitis associated with a rare *Escherichia coli* serotype. *N. Engl. J. Med.* **308**:681–685.
36. Riordan, J. T., S. B. Viswanath, S. D. Manning, and T. S. Whittam. 2008. Genetic differentiation of *Escherichia coli* O157:H7 clades associated with human disease by real-time PCR. *J. Clin. Microbiol.* **46**:2070–2073.
37. Rutherford, K., J. Parkhill, J. Crook, T. Horsnell, P. Rice, M. A. Rajandream, and B. Barrell. 2000. Artemis: sequence visualization and annotation. *Bioinformatics* **16**:944–945.
38. Silagyi, K., S.-H. Kim, Y. Martin Lo, and C.-I. Wei. 2009. Production of biofilm and quorum sensing by *Escherichia coli* O157:H7 and its transfer from contact surfaces to meat, poultry, ready-to-eat deli, and produce products. *Food Microbiol.* **26**:514–519.
39. Tamura, K., J. Dudley, M. Nei, and S. Kumar. 2007. MEGA4: molecular evolutionary genetics analysis (MEGA) software, version 4.0. *Mol. Biol. Evol.* **24**:1596–1599.
40. van Diemen, P. M., F. Dziva, M. P. Stevens, and T. S. Wallis. 2005. Identification of enterohemorrhagic *Escherichia coli* O26:H⁻ genes required for intestinal colonization in calves. *Infect. Immun.* **73**:1735–1743.
41. Wells, J. G., L. D. Shipman, K. D. Greene, E. G. Sowers, J. H. Green, D. N. Cameron, F. P. Downes, M. L. Martin, P. M. Griffin, S. M. Ostroff, et al. 1991. Isolation of *Escherichia coli* serotype O157:H7 and other Shiga-like-toxin-producing *E. coli* from dairy cattle. *J. Clin. Microbiol.* **29**:985–989.
42. Yang, Z., J. Kovar, J. Kim, J. Niefeldt, D. R. Smith, R. A. Moxley, M. E. Olson, P. D. Fey, and A. K. Benson. 2004. Identification of common subpopulations of non-sorbitol-fermenting, β -glucuronidase-negative *Escherichia coli* O157:H7 from bovine production environments and human clinical samples. *Appl. Environ. Microbiol.* **70**:6846–6854.
43. Zhang, W., B. M. Jayarao, and S. J. Knabel. 2004. Multi-virulence-locus sequence typing of *Listeria monocytogenes*. *Appl. Environ. Microbiol.* **70**:913–920.
44. Zhang, W., W. Qi, T. J. Albert, A. S. Motiwala, D. Alland, E. K. Hyttia-Trees, E. M. Ribot, P. I. Fields, T. S. Whittam, and B. Swaminathan. 2006. Probing genomic diversity and evolution of *Escherichia coli* O157 by single nucleotide polymorphisms. *Genome Res.* **16**:757–767.
45. Zhang, Y., C. Laing, M. Steele, K. Ziebell, R. Johnson, A. Benson, E. Taboada, and V. Gannon. 2007. Genome evolution in major *Escherichia coli* O157:H7 lineages. *BMC Genomics* **8**:121.
46. Ziebell, K., M. Steele, Y. Zhang, A. Benson, E. N. Taboada, C. Laing, S. McEwen, B. Ciebin, R. Johnson, and V. Gannon. 2008. Genotypic characterization and prevalence of virulence factors among Canadian *Escherichia coli* O157:H7 strains. *Appl. Environ. Microbiol.* **74**:4314–4323.