Applied and Environmental Microbiology, Mar. 2011, p. 1946–1956 Vol. 77, No. 6
0099-2240/11/$12.00 doi:10.1128/AEM.02625-10
Copyright © 2011, American Society for Microbiology. All Rights Reserved.

Novel Virulence Gene and Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR) Multilocus Sequence Typing Scheme for Subtyping of the Major Serovars of *Salmonella enterica* subsp. *enterica*▽†

Fenyun Liu,[1] Rodolphe Barrangou,[1,2] Peter Gerner-Smidt,[3] Efrain M. Ribot,[3] Stephen J. Knabel,[1] and Edward G. Dudley[1]*

Department of Food Science, The Pennsylvania State University, University Park, Pennsylvania 16802[1]; Danisco USA Inc., 3329 Agriculture Drive, Madison, Wisconsin 53716[2]; and Centers for Disease Control and Prevention, Atlanta, Georgia 30333[3]

Received 8 November 2010/Accepted 21 January 2011

***Salmonella enterica* subsp. *enterica* is the leading cause of bacterial food-borne disease in the United States. Molecular subtyping methods are powerful tools for tracking the farm-to-fork spread of food-borne pathogens during outbreaks. In order to develop a novel multilocus sequence typing (MLST) scheme for subtyping the major serovars of *S. enterica* subsp. *enterica*, the virulence genes *sseL* and *fimH* and clustered regularly interspaced short palindromic repeat (CRISPR) loci were sequenced from 171 clinical isolates from nine *Salmonella* serovars, *Salmonella* serovars Typhimurium, Enteritidis, Newport, Heidelberg, Javiana, I 4,[5],12:i:−, Montevideo, Muenchen, and Saintpaul. The MLST scheme using only virulence genes was congruent with serotyping and identified epidemic clones but could not differentiate outbreaks. The addition of CRISPR sequences dramatically improved discriminatory power by differentiating individual outbreak strains/clones. Of particular note, the present MLST scheme provided better discrimination of *Salmonella* serovar Enteritidis strains than pulsed-field gel electrophoresis (PFGE). This method showed high epidemiologic concordance for all serovars screened except for *Salmonella* serovar Muenchen. In conclusion, the novel MLST scheme described in the present study accurately differentiated outbreak strains/clones of the major serovars of *Salmonella*, and therefore, it shows promise for subtyping this important food-borne pathogen during investigations of outbreaks.**

Salmonella enterica subsp. *enterica* is the leading cause of bacterial food-borne disease in the United States, with approximately 1.4 million human cases each year since 1996, resulting in an estimated 17,000 hospitalizations, more than 500 deaths (9, 49), and a cost estimated as 2.6 billion dollars (U.S. Department of Agriculture Economic Research Service *Salmonella* food-borne illness cost calculator at http://www.ers.usda.gov/Data/FoodborneIllness/salm_Intro.asp). The nine most common human *S. enterica* serovars, *Salmonella* serovars Typhimurium, Enteritidis, Newport, Heidelberg, Javiana, I 4,[5],12:i:−, Montevideo, Muenchen, and Saintpaul, were responsible for more than 60% of human salmonellosis cases based on the Centers for Disease Control and Prevention's (CDC's) annual summary of 2005 (4, 5) and continue to be a major cause of food-borne illness (6, 7, 8, 9, 23). *Salmonella* has been isolated from a broad range of foods (CDC OutbreakNet Foodborne Outbreak Online Database at http://wwwn.cdc.gov/foodborneoutbreaks/), and widespread distribution of these foods makes tracking the transmission of *Salmonella* difficult during investigations of outbreaks. In order to define the routes of transmission of *Salmonella* within the food system, molecular subtyping methods have been employed to distinguish outbreak from non-outbreak-related strains/clones (16).

Serotyping is the most commonly used molecular subtyping method for *Salmonella*. Serotyping distinguishes *Salmonella* based on immunological classification of the H and O antigens (19) and is routinely used for surveillance of this organism. However, serotyping cannot distinguish outbreak strains/clones of the same serotype of *Salmonella.*

Several nucleic acid-based molecular subtyping methods have been used to subtype *Salmonella*, including amplified fragment length polymorphism (AFLP) (18, 32, 36, 42, 46), multiple-locus variable-number tandem-repeat analysis (MLVA) (2, 30, 31, 37), and pulsed-field gel electrophoresis (PFGE) (35). PFGE is currently considered the "gold standard" method for subtyping food-borne pathogens and is the subtyping method used by PulseNet, the molecular surveillance network in the United States and throughout the world to investigate food-borne illnesses and outbreaks (17). To enhance comparability and interpretation, a standardized PFGE protocol and an extensive quality assurance system have been established in PulseNet (17, 35). The main advantage of PFGE is its high discriminatory power (i.e., ability to separate unrelated strains) for subtyping food-borne pathogens, including most of the major serovars of *Salmonella* (27). However, PFGE lacks discriminatory power for highly clonal serovars of *Salmonella*, such as *Salmonella* serovar Enteritidis (17, 50), or clonal phage types like *Salmonella* serovar Typhimurium

* Corresponding author. Mailing address: Department of Food Science, The Pennsylvania State University, 326 Food Science Building, 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 28 January 2011.

DT104 (17). For example, the multistate *Salmonella* serovar Enteritidis outbreak associated with shell eggs in 2010 was caused by the most common XbaI PFGE pattern (JEGX01.0004) for *Salmonella* serovar Enteritidis (7). A similar scenario was also observed recently during the 2010 outbreak associated with Italian-style salami, when the outbreak strain/clone of *Salmonella* serovar Montevideo had the most common PFGE pattern in the PulseNet database (8).

Compared to PFGE, multilocus sequence typing (MLST), which targets nucleotide sequence differences of several DNA loci, has the potential to be a less labor-intensive method. Moreover, DNA sequence data are discrete, unambiguous, highly informative, portable, and reproducible. Although MLST is an attractive subtyping approach, a satisfactory MLST scheme for subtyping multiple serovars of *Salmonella* to the strain level for investigations of outbreaks has yet to be described. MLST schemes targeting housekeeping genes have been developed; however, these schemes usually have much lower discriminatory power than PFGE (14, 24, 29, 46). In order to increase discriminatory power, virulence genes have been included in MLST schemes for subtyping *Salmonella* (15). Virulence genes are commonly under positive, diversifying selection (13) and therefore tend to have more-variable sequences than housekeeping genes (10, 15). MLST schemes using both housekeeping and virulence genes have been used for subtyping *Salmonella* to the serovar level (44) or for discriminating *Salmonella* serovar Typhimurium to the strain level (15). However, with *Salmonella* serovar Enteritidis, one of the most frequent causes of human salmonellosis, comparative genomic analysis (*Salmonella* single-nucleotide polymorphism [SNP] database at http://www.ncbi.nlm.nih.gov/genomes/static/Salmonella_SNPS.html) suggested that virulence genes alone are not discriminatory enough for differentiating strains from different outbreaks (*Salmonella* SNP database). Therefore, additional genome targets with greater sequence diversity than virulence genes are needed in order to create an effective MLST scheme for *Salmonella*.

One of the fastest evolving genetic elements in bacterial genomes are clustered regularly interspaced short palindromic repeats (CRISPRs) (40). CRISPRs have been identified within the genomes of many archaeal and bacterial species, including *Salmonella* (26, 40, 47). CRISPRs encode tandem sequences containing 21- to 47-bp direct repeats (DRs) separated by spacers of similar size (see Fig. S1 in the supplemental material). Spacers are derived from foreign nucleic acids, such as those from phage or plasmids and can protect bacteria from subsequent infection by homologous phage and plasmids (1). Many CRISPR loci are flanked by an AT-rich leader sequence and CRISPR-associated (Cas) genes (see Fig. S1 in the supplemental material) (1, 3, 22). As a bacterial immune system against foreign DNA, CRISPRs evolve rapidly in response to changing phage pools (48). Besides the addition of new spacers, deletion of spacers is also frequently observed (11, 34). Because of the high polymorphism of CRISPRs, they have been successfully used to subtype *Mycobacterium tuberculosis* during investigations of outbreaks (21). CRISPR sequence analysis has also been used to characterize a number of other bacteria, including *Yersinia pestis* (34), serotype M1 group A *Streptococcus* strains (25), and *Campylobacter jejuni* (39).

Two CRISPR loci are found in all *Salmonella* serovars in the CRISPR database (http://crispr.u-psud.fr/crispr/) (47). Generally, the two CRISPR loci have different numbers of repeats/spacers and different sets of spacers. There have been no reports of CRISPRs being used as markers in an MLST scheme for subtyping *Salmonella*. Therefore, the purpose of the present study was to investigate whether MLST based on both virulence genes and CRISPRs can accurately differentiate outbreak strains/clones of the major serovars of *Salmonella*.

MATERIALS AND METHODS

Bacterial isolates and DNA extraction. All 171 *Salmonella enterica* isolates used in this study (Table 1) were from the PulseNet culture collection maintained by the Centers for Disease Control and Prevention (CDC) in Atlanta, GA. This set of isolates represents the 9 serovars most commonly associated with human disease and includes isolates involved in multiple outbreaks, with 2 or 3 isolates per outbreak. In some cases, isolates with different PFGE patterns that were obtained from the same outbreak (had poor epidemiologic concordance by PFGE) were deliberately included. All isolates were previously analyzed by serotyping, and most isolates were analyzed by PFGE by the CDC. Bacterial isolates were stored at −80°C in 20% glycerol. When needed, isolates were grown overnight in tryptic soy broth (TSB) (Difco Laboratories, Becton Dickinson, Sparks, MD) at 37°C. For all isolates, DNA was extracted using the UltraClean microbial DNA extraction kit (Mo Bio Laboratories, Solana Beach, CA) and stored at −20°C before use.

PCR amplification of virulence genes and CRISPRs. *In silico* analysis of 9 publically available whole-genome sequences of *S. enterica* (serovar Agona SL483, GenBank accession no. CP001138; serovar Choleraesuis SC-B67, GenBank accession no. AE017220; serovar Dublin CT_02021853, GenBank accession no. CP001144; serovar Enteritidis P125109, GenBank accession no. AM933172; serovar Gallinarum strain 287191, GenBank accession no. AM933173; serovar Heidelberg SL476, GenBank accession no. CP001120; serovar Newport SL254, GenBank accession no. CP001113; serovar Schwarzengrun CVM19633, GenBank accession no. CP001127; and serovar Typhimurium LT2, GenBank accession no. AE006468) was used to identify 14 virulence genes (*hilA*, *fimH1*, *fimH2*, *pipB*, *sopE*, *sseF*, *sseL*, *sseJ*, *siiA*, *sifB*, *stdA*, *fimA*, *bcfC*, and *phoQ*) (Tables 2 and 3; see Table S1 in the supplemental material) that were present in all genomes but displayed differences in their DNA sequences. Primers for amplifying these genes were designed using Primer 3.0 (http://frodo.wi.mit.edu/primer3/) and were based on the published *Salmonella* serovar Typhimurium LT2 (GenBank accession no. AE006468) genome (Table 3; see Table S1 in the supplemental material). Primers for amplifying CRISPR1 were designed based upon consensus alignments of the published *Salmonella* serovar Typhimurium LT2 (GenBank accession no. AE006468) and serovar Newport strain SL254 genomes (GenBank accession no. CP001113), and the *Salmonella* serovar Javiana strain GA_MM04042433 (GenBank accession no. ABEH00000000) whole-genome shotgun sequence (Table 3). Primers for amplifying CRISPR2 were designed based on the *Salmonella* serovar Typhimurium LT2 genome. All primers annealed to conserved regions located 5′ or 3′ of the CRISPR loci. PCR amplifications were performed using a *Taq* PCR master mix kit (Qiagen Inc., Valencia, CA) and a Mastercycler PCR thermocycler (Eppendorf Scientific, Hamburg, Germany). A 25-μl PCR system contained 12.5 μl of *Taq* PCR 2× master mix, 9.5 μl of PCR-grade water, 1.0 μl of DNA template, 1.0 μl of forward primer (final concentration, 0.4 μM), and 1.0 μl of reverse primer (final concentration, 0.4 μM). A single PCR cycling condition was used for separately amplifying all four markers. PCR was performed as follows: initial denaturation step of 10 min at 94°C; 28 cycles, with 1 cycle consisting of 1 min at 94°C, 1 min at 55°C, and 1 min at 72°C; final extension step of 10 min at 72°C.

DNA sequencing of virulence genes and CRISPRs. After PCR, products for sequencing were treated with 1/20 volume of shrimp alkaline phosphatase (1 U/μl) (USB Corp., Cleveland, OH) and 1/20 volume of exonuclease I (10 U/μl) (USB Corp). The mixture was then incubated at 37°C for 45 min to degrade remaining primers and unincorporated deoxynucleoside triphosphates (dNTPs). After that, the mixture was incubated at 80°C for 15 min to inactivate the added enzymes. PCR products were sent to the Genomics Core Facility at The Pennsylvania State University for sequencing using the ABI data 3730XL DNA analyzer. In order to obtain complete DNA sequences of *fimH* and *sseL*, two more primers targeting the internal regions of these two genes were used together with the forward and reverse primers (Table 3). Both DNA strands of the amplicons were sequenced.

TABLE 1. Outbreak information, PFGE profile, and MLST results for the 171 *Salmonella enterica* isolates analyzed in the present study

CDC code[a]	Source	State	Site or location of isolation	Cluster	PFGE profile		MLST ST[c]
					XbaI	BlnI	
ST29	Water filter	UT	Frog	0909MAJPX-1	JPXX01.0177	JPXA26.0459	T ST1
ST30	Human (stool sample)	MD	Frog	0909MAJPX-1	JPXX01.0177	JPXA26.0459	T ST1
ST31	Human (stool sample)	OH	Frog	0909MAJPX-1	JPXX01.0177	JPXA26.0459	T ST1
ST4	Human (stool sample)	CO	Water	0803COJPX-1c	JPXX01.0002	JPXA26.0002	T ST2
ST5	Water	CO	Water	0803COJPX-1c	JPXX01.0002	JPXA26.0002	T ST2
ST6	Human (stool sample)	OH	Peanut butter	0811MLJPX-1c	JPXX01.0459	JPXA26.0462	T ST3
ST7	Human (stool sample)	OH	Peanut butter	0811MLJPX-1c	JPXX01.1825	JPXA26.0462	T ST3
ST8	Food (peanut butter)	MN	Peanut butter	0811SDCJPX-1c	JPXX01.1818	JPXA26.0462	T ST3
ST9	Stool sample	MA	Raw milk	Outbreak a	JPXX01.0083	JPXA26.0019	T ST4
ST10	Raw milk	MA	Raw milk	Outbreak a	JPXX01.0083	JPXA26.0019	T ST4
ST17	NA[b]	OR	NA	0309ORJPX-1c	JPXX01.0981	JPXA26.0174	T ST4
ST18	NA	OR	NA	0309ORJPX-1c	JPXX01.0981	JPXA26.0174	T ST4
ST11	Stool sample	NM	NA	Outbreak b	JPXX01.0003	JPXA26.0007	T ST5
ST12	Stool sample	NM	NA	Outbreak b	JPXX01.0003	JPXA26.0007	T ST5
ST13	Stool sample	NM	NA	Outbreak b	JPXX01.0003	JPXA26.0008	T ST5
ST26	Human (stool sample)	OR	Snake or mouse	0908ORJPX-1	JPXX01.0003	JPXA26.0003	T ST5
ST27	Human (stool sample)	OR	Snake or mouse	0908ORJPX-1	JPXX01.0003	JPXA26.0003	T ST5
ST28	Animal	OR	Snake or mouse	0908ORJPX-1	JPXX01.0003	JPXA26.0003	T ST5
ST39	Human (stool sample)	VA	Sporadic	Sporadic	JPXX01.0003	JPXA26.0042	T ST5
ST16	Stool sample	MA	Snack	0704WIWWS-c	JPXX01.1037	JPXA26.0333	T ST6
ST19	Stool sample	VT	Snack	0704WIWWS-1c	JPXX01.1037	JPXA26.0333	T ST6
ST20	Stool sample	VT	Snack	0704WIWWS-1c	JPXX01.1037	JPXA26.0333	T ST6
ST32	Human (stool sample)	AR	Day care	0602ARJPX-2c	JPXX01.0010	JPXA26.0233	T ST7
ST33	Human (stool sample)	AR	Day care	0602ARJPX-2c	JPXX01.0010	JPXA26.0233	T ST7
ST34	Human (stool sample)	AR	Day care	0602ARJPX-2c	JPXX01.0010	JPXA26.0233	T ST7
ST40	Human (stool sample)	NY	Sporadic	Sporadic	JPXX01.0003	JPXA26.0042	T ST8
SE1	Human (stool sample)	MN	Stuffed chicken	0603MNJEG-1c	JEGX01.0005	JEGA26.0004	E ST1
SE2	Human (stool sample)	MN	Stuffed chicken	0603MNJEG-1c	JEGX01.0005	JEGA26.0004	E ST1
SE23	Human (stool sample)	MN	NA	0603MNJEG-1c	JEGX01.0005	JEGA26.0004	E ST1
SE18	Human (stool sample)	MN	NA	0803MNJEG-1	JEGX01.0005	JEGA26.0004	E ST1
SE3	Environment	CA	Almonds	Outbreak in 2001	JEGX01.0012	NA	E ST2
SE4	Food (raw almonds)	CA	Almonds	Outbreak in 2001	JEGX01.0012	NA	E ST2
SE5	Environment	CA	Almonds	Outbreak in 2001	JEGX01.0012	NA	E ST2
SE21	Environment	NA	NA	Outbreak in 2001	JEGX01.0013	NA	E ST2
SE25	Environment	NA	Prison	Outbreak in 2001	JEGX01.0013	NA	E ST2
SE6	Human (stool sample)	ME	NA	0612MEJEG-1c	JEGX01.0004	JEGA26.0002	E ST3
SE7	Human (stool sample)	ME	NA	0612MEJEG-1c	JEGX01.0004	JEGA26.0002	E ST3
SE26	Human (stool sample)	CO	NA	NA	JEGX01.0004	JEGA26.0002	E ST3
SE31	Human (stool sample)	CO	NA	NA	JEGX01.0004	JEGA26.0002	E ST3
SE24	Human (stool sample)	WV	NA	NA	JEGX01.0004	JEGA26.0002	E ST3
SE8	Human (stool sample)	PA	Egg	0801PAJEG-1	JEGX01.0004	JEGA26.0002	E ST4
SE9	Human (stool sample)	PA	Egg	0801PAJEG-1	JEGX01.0004	JEGA26.0002	E ST4
SE15	Human (stool sample)	PA	NA	0801PAJEG-1	JEGX01.0004	JEGA26.0002	E ST4
SE34	Human (stool sample)	CT	NA	NA	JEGX01.0004	JEGA26.0002	E ST4
SE11	Human (stool sample)	GA	Hospital eggs	0505GAJEG-1c	JEGX01.0018	JEGA26.0005	E ST4
SE10	Human (stool sample)	GA	Hospital eggs	0505GAJEG-1c	JEGX01.0034	JEGA26.0005	E ST5
SE12	NA	ME	NA	0612MEJEG-1c	JEGX01.0004	JEGA26.0002	E ST6
SE13	Human (stool sample)	ME	NA	0612MEJEG-1c	JEGX01.0004	JEGA26.0002	E ST6
SE14	Human (stool sample)	ME	NA	0612MEJEG-1c	JEGX01.0004	JEGA26.0002	E ST7
SE16	Human (stool sample)	GA	NA	0506GAJEG-1c	JEGX01.0004	JEGA26.0002	E ST8
SE19	Human (stool sample)	GA	NA	0506GAJEG-1c	JEGX01.0004	JEGA26.0002	E ST8
SE30	Human (stool sample)	GA	Prison	0506GAJEG-1c	JEGX01.0004	JEGA26.0002	E ST8
SE22	Human (stool sample)	OR	NA	0509ORJEG-1c	JEGX01.0004	JEGA26.0025	E ST8
SE27	Human (stool sample)	OR	NA	0509ORJEG-1c	JEGX01.0004	JEGA26.0025	E ST8
SE28	Human	SC	NA	0504SCJEG-1c	JEGX01.0004	JEGA26.0002	E ST8
SE29	Human (stool sample)	ID	NA	0504CAOCJEG-1c	JEGX01.0004	JEGA26.0002	E ST8
SE17	NA	OH	Frozen chicken	Outbreak in 2005 (2005-28-076)	JEGX01.0005	JEGA26.0004	E ST9
SE20	NA	OH	NA	Outbreak in 2005 (2005-28-076)	JEGX01.0005	JEGA26.0004	E ST9
SE32	Human (stool sample)	MI	NA	0708MIJEG-1c	JEGX01.0005	JEGA26.0004	E ST9
SE33	Human (stool sample)	MI	NA	0708MIJEG-1c	JEGX01.0005	JEGA26.0004	E ST9
SN1	NA	IL	NA	NA	JJPX01.0014	NA	N ST1
SN2	NA	IL	NA	NA	JJPX01.0014	NA	N ST1
SN3	NA	NA	NA	0509NHJJP-1c.	JJPX01.0061	JJPA26.0021	N ST2
SN4	NA	NA	NA	0509NHJJP-1c.	JJPX01.0061	JJPA26.0021	N ST2
SN5	NA	NA	NA	NA	JJPX01.0001	NA	N ST3
SN6	NA	NA	NA	NA	JJPX01.0001	NA	N ST3

Continued on following page

TABLE 1—*Continued*

CDC code[a]	Source[b]	State	Site or location of isolation	Cluster	PFGE profile		MLST ST[c]
					XbaI	BlnI	
SN7	Human (stool sample)	CA	NA	0710CAJJP-1c	JJPX01.0422	JJPA26.0196	N ST4
SN8	Human (stool sample)	CA	NA	0710CAJJP-1c	JJPX01.0422	JJPA26.0196	N ST4
SN11	Human (stool sample)	SD	NA	0712SDJJP-1c	JJPX01.0654	JJPA26.0208	N ST4
SN12	Human (stool sample)	SD	NA	0712SDJJP-1c	JJPX01.0654	JJPA26.0208	N ST4
SN9	Human (stool sample)	AZ	NA	0802AZJJP-1c	JJPX01.0696	JJPA26.0212	N ST5
SN10	Human (stool sample)	AZ	NA	0802AZJJP-1c	JJPX01.0438	JJPA26.0212	N ST5
SN13	Human (stool sample)	GA	NA	0711GAJJP-1c	JJPX01.1319	JJPA26.0542	N ST6
SN14	Human (stool sample)	GA	NA	0711GAJJP-1c	JJPX01.1319	JJPA26.0542	N ST6
SN15	Human (stool sample)	GA	NA	0711GAJJP-1c	JJPX01.1319	JJPA26.0542	N ST6
SH1	Human	DE	Cruise ship	0607NYJF6-1c	JF6X01.0022	NA	H ST1
SH2	Human	NY	Cruise ship	0607NYJF6-1c	JF6X01.0022	NA	H ST1
SH3	Human	NY	Cruise ship	0607NYJF6-1c	JF6X01.0022	NA	H ST1
SH8	Human	IL	Hummus	0707ILJF6-1c	JF6X01.0032	JF6A26.0076	H ST1
SH9	Human	IL	Hummus	0707ILJF6-1c	JF6X01.0032	JF6A26.0076	H ST1
SH10	Human	IL	Hummus	0707ILJF6-1c	JF6X01.0032	JF6A26.0076	H ST1
SH11	Human	IL	Hummus	0707ILJF6-1c	JF6X01.0032	JF6A26.0076	H ST1
SH16	NA	NA	Sporadic	Sporadic	JF6X01.0122	NA	H ST1
SH17	NA	NA	Sporadic	Sporadic	JF6X01.0022	NA	H ST1
SH18	Human	NA	NA	0704AZJPX-1c	JF6X01.0022	NA	H ST1
SH4	Human	PA	Religious camp	0607PAJF6-1c	JF6X01.0022	NA	H ST2
SH5	Human	PA	Religious camp	0607PAJF6-1c	JF6X01.0022	NA	H ST2
SH6	Human	PA	Religious camp	0607PAJF6-1c	JF6X01.0022	NA	H ST2
SH7	Human	PA	Religious camp	0607PAJF6-1c	JF6X01.0022	NA	H ST2
SH15	NA	NA	Sporadic	Sporadic	JF6X01.0051	NA	H ST2
SH12	Human	TN	NA	0702TNJF6-1c	JF6X01.0032	JF6A26.0076	H ST3
SH13	Human	TN	NA	0702TNJF6-1c	JF6X01.0032	JF6A26.0076	H ST3
SH14	NA	NA	Sporadic	Sporadic	JF6X01.0135	NA	H ST4
SH19	Human	NA	NA	0704AZJPX-1c	JF6X01.0022	NA	H ST5
SH20	Human	NA	NA	0704AZJPX-1c	JF6X01.0022	NA	H ST6
SJ1	NA	AL	NA	NA	JGGX01.0012	NA	J ST1
SJ5	NA	AR	NA	NA	JGGX01.0012	NA	J ST1
SJ13	NA	LA	NA	NA	NA	NA	J ST1
SJ15	NA		Outbreak	NA	JGGX01.0036	JGGA26.0017	J ST1
SJ2	NA	TX	NA	NA	JGGX01.0213	NA	J ST2
SJ3	NA	LA	NA	NA	NA	NA	J ST3
SJ8	NA	LA	NA	NA	NA	NA	J ST3
SJ4	NA	TX	NA	NA	NA	NA	J ST4
SJ6	NA	AR	NA	NA	JGGX01.0179	NA	J ST5
SJ9	NA	AR	NA	NA	JGGX01.1226	NA	J ST5
SJ7	NA	TX	NA	NA	JGGX01.1525	NA	J ST6
SJ10	NA	HU	NA	NA	NA	NA	J ST7
SJ11	NA	MD	NA	NA	JGGX01.0362	NA	J ST8
SJ12	NA	IL	NA	NA	JGGX01.1352	NA	J ST9
SJ14	NA	NV	NA	NA	NA	NA	J ST10
ST1[d]	Stool sample	CA	Turkey potpie	0706PAJPX-1c	JPXX01.0206	JPXA26.0180	I ST1
ST2[d]	Stool sample	GA	Turkey potpie	0706PAJPX-1c	JPXX01.0206	JPXA26.0180	I ST1
ST3[d]	Food (turkey potpie)	WI	Turkey potpie	0706PAJPX-1c	JPXX01.0206	JPXA26.0180	I ST1
ST14[d]	Stool sample	IN	NA	0607INjpx-1c	JPXX01.0621	JPXA26.0160	I ST1
ST15[d]	Stool sample	IN	NA	0607INjpx-1c	JPXX01.0621	JPXA26.0160	I ST1
ST21[d]	Human (stool sample)	OH	Snake	0806OHJPX-1c	JPXX01.1596	JPXA26.0491	I ST2
ST22[d]	Human (stool sample)	OH	Snake	0806OHJPX-1c	JPXX01.1596	JPXA26.0491	I ST2
ST23[d]	Human (stool sample)	OH	Snake	0806OHJPX-1c	JPXX01.1596	JPXA26.0491	I ST2
ST24[d]	Food (egg wash)	ME	Egg	0404PAJPX-1c	JPXX01.0621	JPXA26.0057	I ST3
ST25[d]	NA	VT	Egg	0404PAJPX-1c	JPXX01.0621	JPXA26.0057	I ST3
ST35[d]	Human (stool sample)	OH	Sporadic	Sporadic	JPXX01.0621	JPXA26.0055	I ST4
ST36[d]	Human (stool sample)	MA	Sporadic	Sporadic	JPXX01.1212	JPXA26.0108	I ST4
ST37[d]	Human (stool sample)	MO	Sporadic	Sporadic	JPXX01.0206	JPXA26.0380	I ST4
SMvo1	Blood sample	TX	NA	NA	NA	NA	Mvo ST1
SMvo2	Stool sample	TX	NA	NA	NA	NA	Mvo ST2
SMvo7	Human	MD	NA	NA	JIXX01.0524	NA	Mvo ST3
SMvo3	Human (rectal swab sample)	AZ	Raw chicken	0807AZJIX-1c	JIXX01.1014	NA	Mvo ST3
SMvo8	Human (stool sample)	AZ	Raw chicken	0807AZJIX-1c	JIXX01.0126	JIXA26.0012	Mvo ST3
SMvo9	Human (swab sample)	AZ	Raw chicken	0807AZJIX-1c	JIXX01.0126	JIXA26.0012	Mvo ST3
SMvo10	Human (stool sample)	AZ	Raw chicken	0807AZJIX-1c	JIXX01.0126	JIXA26.0012	Mvo ST3
SMvo11	Human (stool sample)	UT	Salami or pepper	0908ORJIX-1	JIXX01.0011	JIXA26.0012	Mvo ST3

Continued on following page

TABLE 1—*Continued*

CDC code[a]	Source[b]	State	Site or location of isolation	Cluster	PFGE profile		MLST ST[c]
					XbaI	BlnI	
SMvo12	Human (urine sample)	OR	Salami or pepper	0908ORJIX-1	JIXX01.0011	JIXA26.0012	Mvo ST3
SMvo13	Human (stool sample)	AZ	Salami or pepper	0908ORJIX-1	JIXX01.0011	NA	Mvo ST3
SMvo15	Human (stool sample)	TN	Salami or pepper	0908ORJIX-1	JIXX01.0011	NA	Mvo ST3
SMvo14	NA	AZ	NA	NA	NA	NA	Mvo ST3
SMvo4	NA	TX	NA	NA	JIXX01.0388	NA	Mvo ST4
SMvo5	Human (stool sample)	TX	NA	NA	JIXX01.0875	NA	Mvo ST5
SMvo6	Human (stool sample)	TN	NA	NA	JIXX01.1005	NA	Mvo ST6
SMcn1	NA	TX	NA	Outbreak c	JJPX01.0014	NA	Mcn ST1
SMcn2	NA	NY	NA	Outbreak c	JJPX01.0014	NA	Mcn ST2
SMcn3	Human (stool sample)	LA	NA	0509NHJJP-1c	JJPX01.0061	JJPA26.0021	Mcn ST3
SMcn4	NA	TX	NA	0509NHJJP-1c	JJPX01.0061	JJPA26.0021	Mcn ST4
SMcn5	NA	TX	NA	NA	NA	NA	Mcn ST5
SMcn6	Human (stool sample)	TX	NA	NA	NA	NA	Mcn ST6
SMcn7	NA	TX	NA	0710CAJJP-1c	JJPX01.0422	JJPA26.0196	Mcn ST7
SMcn8	NA	TX	NA	0710CAJJP-1c	JJPX01.0422	JJPA26.0196	Mcn ST8
SMcn9	Human (stool sample)	TX	NA	0802AZJJP-1c	JJPX01.0696	JJPA26.0212	Mcn ST9
SMcn10	NA	TX	NA	0802AZJJP-1c	JJPX01.0438	JJPA26.0212	Mcn ST10
SMcn11	Human (stool sample)	TX	NA	0712SDJJP-1c	JJPX01.0654	JJPA26.0208	Mcn ST11
SMcn12	Human	MD	NA	0712SDJJP-1c	JJPX01.0654	JJPA26.0208	Mcn ST12
SMcn13	NA	OR	Orange juice	0711GAJJP-1c	JJPX01.1319	JJPA26.0542	Mcn ST13
SMcn15	NA	WA	Orange juice	0711GAJJP-1c	JJPX01.1319	JJPA26.0542	Mcn ST13
SMcn14	NA	WA	Orange juice	0711GAJJP-1c	JJPX01.1319	JJPA26.0542	Mcn ST14
SS10	Human	MA	NA	0806MAJN6-1c	JN6X01.0034	JN6A26.0038	S ST1
SS11	Human	MA	NA	0806MAJN6-1c	JN6X01.0034	JN6A26.0038	S ST1
SS12	Human	MA	NA	0806MAJN6-1c	JN6X01.0034	JN6A26.0038	S ST1
SS6	Human	NE	Sprouts	0902NEJN6-1	JN6X01.0072	NA	S ST2
SS7	Human	NE	Sprouts	0902NEJN6-1	JN6X01.0072	NA	S ST2
SS8	Human	NE	Sprouts	0902NEJN6-1	JN6X01.0072	NA	S ST2
SS9	Human	NE	Sprouts	0902NEJN6-1	JN6X01.0072	NA	S ST2
SS1	NA	MN	Jalapeños	0805NMJN6-1c	JN6X01.0048	JN6A26.0019	S ST3
SS2	Human	TX	Jalapeños	0805NMJN6-1c	JN6X01.0048	JN6A26.0019	S ST3
SS3	Human	NM	Jalapeños	0805NMJN6-1c	JN6X01.0048	JN6A26.0019	S ST3
SS4	Human	AZ	Jalapeños	0805NMJN6-1c	JN6X01.0048	JN6A26.0019	S ST3
SS18	NA	NE	Sporadic	Sporadic	JN6X01.0622	NA	S ST3
SS19	NA	TX	Sporadic	Sporadic	JN6X01.0067	JN6A26.0001	S ST3
SS16	NA	CA	Sporadic	Sporadic	JN6A26.0026	NA	S ST4
SS13	NA	CA	NA	0807LACJN6-1c	JN6X01.0021	JN6A26.0019	S ST5
SS14	NA	CA	NA	0807LACJN6-1c	JN6X01.0021	JN6A26.0019	S ST5
SS15	NA	MD	Sporadic	Sporadic	JN6X01.0170	NA	S ST5
SS20	NA	NV	Sporadic	Sporadic	JN6X01.0623	JN6A26.0047	S ST6

[a] The CDC code used for isolates follows: ST, *Salmonella* serovar Typhimurium (ST29 to ST31 are isolates of *Salmonella* serovar Typhimurium subsp. Copenhagen); SE, *Salmonella* serovar Enteritidis; SN, *Salmonella* serovar Newport; SH, *Salmonella* serovar Heidelberg; SJ, *Salmonella* serovar Javiana; SI, *Salmonella* serovar I 4,[5],12:i:−; SMvo, *Salmonella* serovar Montevideo; SMcn, *Salmonella* serovar Muenchen; SS, *Salmonella* serovar Saintpaul.

[b] NA, not available.

[c] Sequence types (STs) based upon the combination of *fimH1*, *sseL*, CRISPR1, and CRISPR2 are abbreviated as follows: T, *Salmonella* serovar Typhimurium; E, *Salmonella* serovar Enteritidis; N, *Salmonella* serovar Newport; H, *Salmonella* serovar Heidelberg; J, *Salmonella* serovar Javiana; Mvo, *Salmonella* serovar Montevideo; Mcn, *Salmonella* serovar Muenchen; S, *Salmonella* serovar Saintpaul. For instance, T ST1 stands for *Salmonella* serovar Typhimurium sequence type 1.

[d] ST1 to ST3, ST14 and ST15, ST21 to ST25, and ST35 to ST37 are isolates of *Salmonella* serovar I 4,[5],12;i:−.

TABLE 2. Size, function, and nucleotide location of the four markers targeted in the present study

Marker	Size (bp)	Function(s)	Nucleotide location in *Salmonella* serovar Typhimurium LT2
fimH	1,008	Host-cell-specific recognition	28425–29432
sseL	954	Inflammation and macrophage killing	2394795–2395748
CRISPR1	122–854[a]	Defense against phage	3076611–3077006
CRISPR2	183–1,525[a]	Defense against phage	3094279–3096260

[a] The length of CRISPRs varied because the number of repeats/spacers changed among the different strains analyzed.

Sequence analysis and sequence type assignment. Virulence gene sequences were aligned, and single-nucleotide polymorphisms (SNPs) were identified using MEGA 4.0 (43). For CRISPR1 and CRISPR2, analyses of the spacer arrangements were performed using CRISPRcompar (20), and spacers were visualized by the method of Deveau et al. (11). Different allelic types (ATs) (sequences with at least one-nucleotide difference or one-spacer difference in the case of CRISPRs) were assigned arbitrary numbers. The combination of 4 alleles (*fimH*, *sseL*, CRISPR1, and CRISPR2) determined its allelic profile, and each unique allelic profile was designated a unique sequence type (ST). The epidemiological relationships of the strains were kept from the investigators until the data had been analyzed and sequence types assigned.

Calculation of epidemiologic concordance. Epidemiologic concordance (*E*) was calculated using the equation developed by the European Study Group on Epidemiologic Markers (41).

Cluster analysis. Cluster analyses were performed based on allelic profile data, and results were visualized using the tree drawing tool on PubMLST (www.pubmlst.org). CRISPR1 and CRISPR2 might be genetically linked due to

TABLE 3. Primers used to amplify and sequence the four MLST markers

Marker	Primer sequence (5′–3′)	Description and function
fimH	CGTCGTCATAAAAGGAAAAA	Forward primer for both amplification and sequencing
	GAACAAAACACAACCAATAGC	Reverse primer for both amplification and sequencing
	CTCGCCAGACAATGTTTACT	Reverse primer for sequencing internal region
	CATTCACTTCGCAGTTTTG	Forward primer for sequencing internal region
sseL	AGGAAACAGAGCAAAATGAA	Forward primer for both amplification and sequencing
	TAAATTCTTCGCAGAGCATC	Reverse primer for both amplification and sequencing
	GGAGTTGAAAATCTTTGGTG	Reverse primer for sequencing internal region
	TTTACCGAGAGAAAAGGTGA	Forward primer for sequencing internal region
CRISPR1	GATGTAGTGCGGATAATGCT	Forward primer for both amplification and sequencing
	GGTTTCTTTTCTTCCTGTTG	Reverse primer for both amplification and sequencing[a]
	GATGATATGGCAACAGGTTT	Reverse primer for both amplification and sequencing[a]
	TATTGACTGCGATGAGATGA	Reverse primer for both amplification and sequencing[b]
CRISPR2	ACCAGCCATTACTGGTACAC	Forward primer for both amplification and sequencing
	ATTGTTGCGATTATGTTGGT	Reverse primer for both amplification and sequencing

[a] The two reverse primers (reverse 1 and reverse 2) of CRISPR1 were added together with the forward primer to amplify CRISPR1 in all serovars except *Salmonella* serovar Javiana.

[b] The reverse primer for SJ (*Salmonella* serovar Javiana) was needed for amplification and sequencing of CRISPR1 in *Salmonella* serovar Javiana isolates.

their proximity in the genome (47); therefore, they were combined into one allele to reduce their weight in the cluster analysis (Fig. 1c).

Statistical analysis. The standard deviations of the average numbers of spacers in CRISPR1 and CRISPR2 were calculated using Microsoft Excel.

Nucleotide sequence accession numbers. DNA sequences of the four genetic MLST markers were deposited in GenBank under accession numbers HQ329797 to HQ329931.

RESULTS

Virulence genes alone provide limited discrimination of *Salmonella* isolates. We began this study by sequencing 14 virulence genes (*fimH*, *sseL*, *hilA*, *fimH2*, *pipB*, *sopE*, *sseF*, *sseJ*, *siiA*, *sifB*, *stdA*, *fimA*, *bcfC*, and *phoQ*) from 20 *Salmonella* serovar Typhimurium, 15 *Salmonella* serovar Newport, and 15 *Salmonella* serovar Enteritidis isolates. Two virulence genes, *fimH* and *sseL*, were found to provide discrimination equal to the combined discrimination of all 14 virulence genes (data not shown); therefore, the other 12 virulence genes were excluded from the rest of the study. Virulence genes *fimH* and *sseL* were sequenced from the remaining isolates, and the total number of allelic types was 17 for *fimH* and 16 for *sseL* (Table 4). Only epidemiologically unrelated strains were included in the calculation of polymorphic sites. The total number (percentage of polymorphic sites) for *fimH* was 48 (4.76%), and for *sseL*, it was 69 (7.23%) (Table 5). Within each serovar, the percentage of polymorphic sites in *fimH* ranged from 0% to 1.79%; for *sseL*, the percentage of polymorphic sites ranged from 0% to 3.88%. For both *fimH* and *sseL*, less polymorphism was observed for *Salmonella* serovars Typhimurium, Enteritidis, Heidelberg, Javiana, and I 4,[5],12:i:− than for *Salmonella* serovars Newport, Montevideo, Muenchen, and Saintpaul (Table 5). Sequences of *sseL* were especially conserved in *Salmonella* serovars Typhimurium, Heidelberg, Javiana, and I 4,[5],12:i:−, with no SNPs observed within each serovar. For all serovars, a total of 39 polymorphic sites in *sseL* were nonsynonymous, and 13 polymorphic sites in *fimH* were nonsynonymous (Table 5).

Addition of CRISPR1 and CRISPR2 to the MLST scheme significantly increases discriminatory power. Since the discrimination provided by virulence genes was limited (separation to outbreak level was not achieved), the addition of CRISPR1 and CRISPR2 to the MLST scheme was investigated. The total numbers of unique spacers in CRISPR1 and CRISPR2 for all 171 isolates analyzed were 166 and 182, respectively (Table 6; see Fig. S2 in the supplemental material). Repeat sequences of the two CRISPRs were generally conserved as shown by the typical repeat in Table S2 in the supplemental material, however, SNPs were sometimes observed and we define these as "repeat variants" (see Table S2 in the supplemental material). The number of spacers in CRISPR1 ranged from 3 to 24, while the number of spacers in CRISPR2 ranged from 2 to 25 (Table 6; see Fig. S2 in the supplemental material). CRISPR2 had more spacers than CRISPR1 for all serovars except serovar Muenchen (Table 6 and Fig. S2).

The number of allelic types for CRISPR1 (44 allelic types) and CRISPR2 (51 allelic types) were significantly greater than those for virulence genes (Table 4). In total, there were 61 sequence types based on both virulence genes and CRISPRs for all 158 isolates that were epidemiologically unrelated (Table 4). An equal number of allelic types was observed in both CRISPR1 and CRISPR2 for *Salmonella* serovars Javiana and Montevideo (Table 4). However, for *Salmonella* serovars Typhimurium, Enteritidis, Newport, Heidelberg, and Saintpaul, CRISPR2 yielded more allelic types than CRISPR1. In contrast, for *Salmonella* serovar Muenchen, CRISPR1 yielded more allelic types than CRISPR2 (Table 4).

CRISPR sequences allow discrimination of isolates within *Salmonella* serovars. Cluster diagrams based on allelic profiles were constructed using only the two virulence genes (Fig. 1a), only CRISPR1 and CRISPR2 (Fig. 1b), and using virulence genes combined with CRISPR (Fig. 1c). Virulence genes alone were effective at separating isolates of different serovars, while the addition of CRISPR1 and CRISPR2 provided additional discrimination between isolates within the same serovar (compare Fig. 1a to Fig. 1c). CRISPR sequencing alone provided the same level of discrimination as the combination of

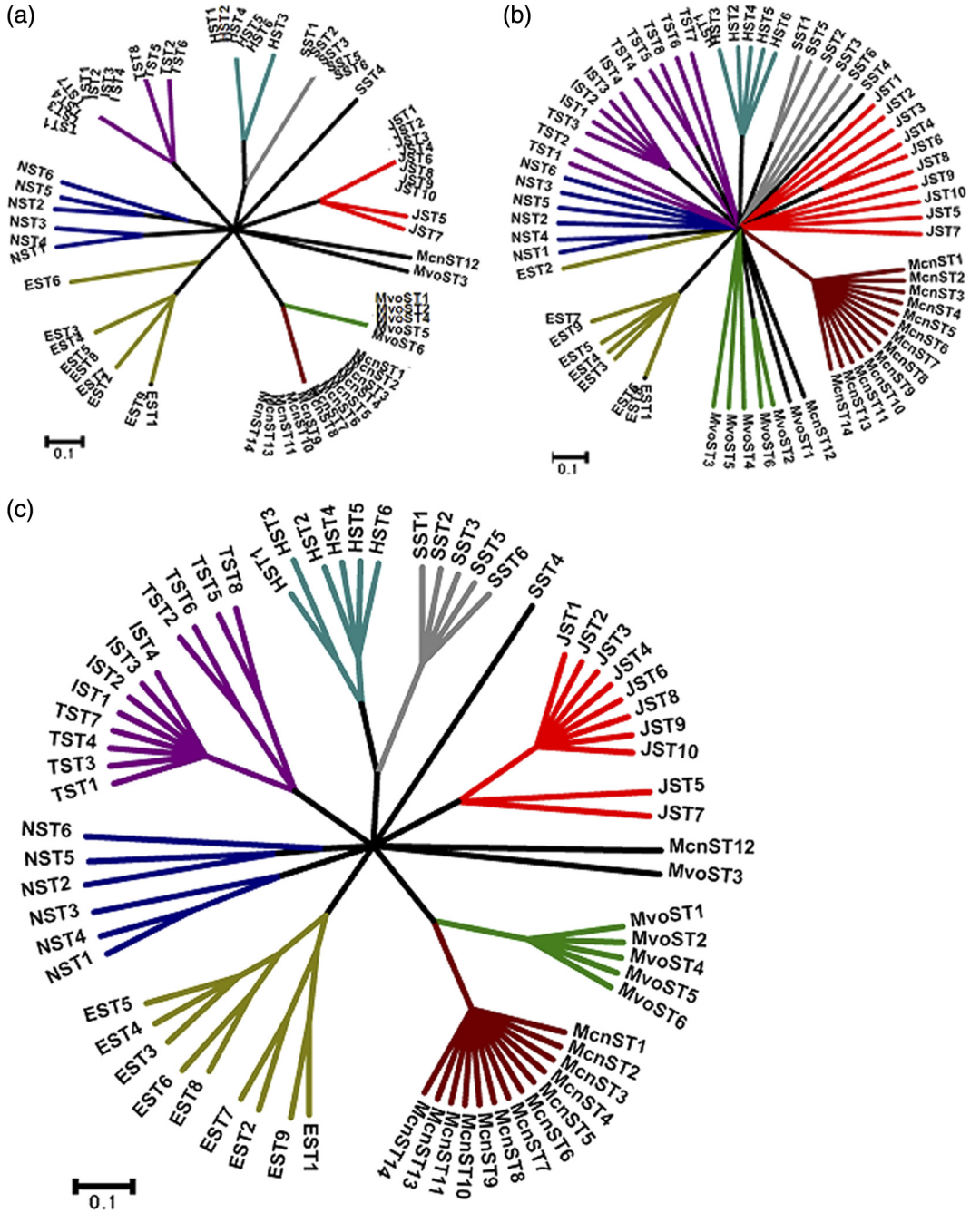

FIG. 1. (a) Cluster diagram based on *fimH* and *sseL*. (b) Cluster diagram based on CRISPR1 and CRISPR2. (c) Cluster diagram based on *fimH*, *sseL*, and CRISPRs (combined allele of CRISPR1 and CRISPR2). Sequence types are abbreviated ST (e.g., ST1). *Salmonella* serovars are shown before the sequence type as follows: T, Typhimurium; E, Enteritidis; N, Newport; H, Heidelberg; J, Javiana; I, I 4,[5],12:i:−; Mvo, Montevideo; Mcn, Muenchen; and S, Saintpaul. In panels b and c, CRISPR1 and CRISPR2 were combined into one allele for the cluster analysis because CRISPR1 and CRISPR2 are genetically linked (47). The scale bar indicates UPGMA linkage distance.

CRISPRs and virulence genes for all serovars except *Salmonella* serovar Enteriditis and *Salmonella* serovar Heidelberg (compare Fig. 1b to Fig. 1c). MLST results showed high congruence with serotypes of *Salmonella*, as isolates of the same serovars typically occupied the same branch of the cluster diagram (Fig. 1c). The three exceptions to this were strains SST4, McnST12, and MvoST3, which occupied unique branches. MLST also did not separate all isolates of the related *Salmonella* serovars Typhimurium and I 4,[5],12:i:−.

MLST discriminates between *Salmonella* serovar Enteritidis strains with identical pulsotypes. Inclusion of CRISPR in the present MLST scheme added to the discrimination provided by PFGE for outbreak isolates of *Salmonella* serovar Enteritidis (Tables 1 and 4). Most isolates of *Salmonella* serovar Enter-

TABLE 4. Number of isolates, allelic types, sequence types, and PFGE patterns in each *Salmonella* serovar[a]

Salmonella serovar	No. of isolates	No. of allelic types				No. of MLST STs	No. of PFGE patterns
		fimH	*sseL*	CRISPR1	CRISPR2		
Typhimurium	24	3	1	7	8	8	11
Enteritidis	32	2	3	2	5	8	5
Newport	14	3	4	4	6	6	7
Heidelberg	19	2	1	1	4	5	5
Javiana	15	3	1	10	10	10	8
I 4,[5],12:i:−	13	1	1	1	4	4	7
Montevideo	14	2	2	6	6	6	6
Muenchen	9	2	2	8	2	8	6
Saintpaul	18	2	2	5	6	6	10
Total	158	17[b]	16	44	51	61	65

[a] This table includes only isolates that are epidemiologically unrelated.
[b] The total number of allelic types for *fimH* does not equal the sum of allelic types in each serovar, because the same allelic type was sometimes present in more than one serovar.

TABLE 5. Allelic polymorphisms and nucleotide substitutions in the nucleotide sequences of *fimH* and *sseL*[a]

Gene	*Salmonella* serovar	No. of polymorphic sites	% of polymorphic sites	No. of synonymous substitutions	No. of nonsynonymous substitutions
fimH	Typhimurium	2	0.2	1	1
	Enteritidis	1	0.1	0	1
	Newport	10	0.99	6	4
	Heidelberg	1	0.1	1	0
	Javiana	2	0.2	0	2
	I 4,[5],12:i:−	0	0	0	0
	Montevideo	13	1.29	10	3
	Muenchen	16	1.59	13	3
	Saintpaul	18	1.79	14	4
	Total	48	4.76	35	13
sseL	Typhimurium	0	0	0	0
	Enteritidis	2	0.21	1	1
	Newport	18	1.89	8	10
	Heidelberg	0	0	0	0
	Javiana	0	0	0	0
	I 4,[5],12:i:−	0	0	0	0
	Montevideo	10	1.05	4	6
	Muenchen	6	0.63	3	3
	Saintpaul	37	3.88	15	22
	Total	69	7.23	30	39

[a] This table includes only isolates that are epidemiologically unrelated.

itidis (25 out of 34) had either the XbaI and BlnI PFGE profile JEGX01.0005 and JEGA26.0004 or JEGX01.0004 and JEGA26.0002 (Table 1). Isolates SE1, SE2, SE23, SE18, SE17, SE20, SE32, and SE33 (CDC code for isolates explained in Table 1, footnote *a*) had the same PFGE profile (JEGX01.0005 and JEGA26.0004) but had two MLST sequence types (E ST1 and E ST 9; MLST sequence types explained in Table 1, footnote *c*) (Table 1). Also, the PFGE profile (JEGX01.0004 and JEGA26.0002), which included isolates SE6, SE7, SE8, SE9, SE15, SE16, SE19, SE30, SE12, SE13, SE14, SE26, SE31, SE28, SE29, SE24, and SE34, were further separated into five sequence types (E ST3, E ST4, E ST6, E ST7, and E ST8) by MLST (Table 1). We did not calculate the discriminatory power (27) of PFGE and MLST, because the isolates used were not randomly selected but were biased toward outbreak strains that showed poor epidemiologic concordance by PFGE.

PFGE provided better separation than MLST for five *Salmonella* serovars screened. For some *Salmonella* serovars (*Salmonella* serovars Newport, Typhimurium, I 4,[5],12:i:−, Montevideo, and Saintpaul), PFGE provided greater separation than MLST among strains associated with different outbreaks. For example, PFGE separated *Salmonella* serovar I 4,[5],12:i:− isolates (ST1, ST2, and ST3) of an outbreak linked to consumption of turkey pot pies (cluster 0706PAJPX-1c) from isolates (ST14 and ST15) of cluster 0607INjpx-1c, while these isolates could not be distinguished by MLST (Table 1). PFGE also distinguished *Salmonella* serovar Typhimurium isolates from an outbreak linked to raw milk consumption (designated "outbreak a" in Table 1) and outbreak cluster 0309ORJPX-1c (Table 1). Also, in contrast to MLST, PFGE was able to discriminate the outbreak linked to raw chicken (cluster 0807AZJIX-1c) from the outbreak linked to salami/pepper (cluster 0908ORJIX-1) of *Salmonella* serovar Montevideo (Table 1). Multiple PFGE patterns were seen among *Salmonella* serovar Newport isolates from MLST sequence types N ST4 and N ST5. For *Salmonella* serovar Saintpaul, both methods allowed accurate separation and identification of all outbreaks due to this serovar, although PFGE provided better separation of sporadic isolates SS18, SS19, and SS15 from outbreak isolates (Table 1).

MLST and PFGE provided complementary information for *Salmonella* serovar Heidelberg. For *Salmonella* serovar Heidelberg, the most accurate outbreak identification was achieved by combining MLST and PFGE. MLST provided separation for the isolates from an outbreak on a cruise ship (cluster 0607NYJF6-1c) and the isolates from an outbreak in a religious camp (cluster 0607PAJF6-1c), which could not be distinguished by PFGE (Table 1). Similarly, MLST distinguished between isolates from an outbreak linked to hummus (cluster JF6X01.0032) and isolates from an outbreak (cluster 0702TNJF6-1c), which had the same pulsotypes in Table 1. However, PFGE separated the isolates from the outbreak on a cruise ship from the isolates from the outbreak linked to hummus, which were indistinguishable by MLST (Table 1).

MLST has high epidemiologic concordance for most *Salmonella* serovars. Values of epidemiologic concordance of MLST and PFGE for each serovar were calculated (Table 7), except for the *Salmonella* serovar Javiana which did not contain any isolates with a defined PulseNet cluster code. Epidemiologic concordance values were calculated based on isolates from well-defined outbreaks (isolates with the same cluster codes), so sporadic isolates and isolates without cluster code designations were excluded. It is important to note that many outbreak isolates included in this study were deliberately chosen because they displayed poor epidemiologic concordance by PFGE. For instance, isolates ST6, ST7, and ST8 were all associated with a 2008 outbreak linked to peanut butter, but each of these isolates had a distinct PFGE pattern (Table 1). MLST showed high epidemiologic concordance (epidemiologic concordance between 0.88 and 1.0) for subtyping all serovars included in this study, except for *Salmonella* serovar Muenchen (epidemiologic concordance of 0.39) (Table 7). On the basis of the limited number of strains analyzed in the present study, MLST showed higher epidemiologic concordance than PFGE for *Salmonella* serovars Enteritidis, Typhimurium, Newport, and Montevideo,

TABLE 6. Analysis of CRISPR spacers in different *Salmonella* serovars[a]

Salmonella serovar	No. of unique spacers		Avg no. of spacers ± SD		Minimum no. of spacers		Maximum no. of spacers	
	CRISPR1	CRISPR2	CRISPR1	CRISPR2	CRISPR1	CRISPR2	CRISPR1	CRISPR2
Typhimurium	26	34	11.4 ± 4.0	19.6 ± 6.8	3	4	14	25
Enteritidis	9	10	8.5 ± 0.6	8.8 ± 1.6	8	7	9	10
Newport	31	43	11.3 ± 4.9	16.3 ± 3.4	4	10	14	19
Heidelberg	10	18	10.0 ± 0.0	12.6 ± 2.7	10	10	10	17
Javiana	9	16	6.4 ± 2.0	9.4 ± 4.0	4	2	9	14
I 4,[5],12:i:−	13	23	13 ± 0	24 ± 1	13	13	23	25
Montevideo	38	40	13.2 ± 5.6	17.7 ± 3.0	9	14	24	22
Muenchen	34	5	12.8 ± 5.0	2.5 ± 0.7	6	2	20	3
Saintpaul	35	33	12.2 ± 1.3	16.5 ± 5.6	11	7	14	23
Total[b]	166	182	10.8 ± 4.5	14.4 ± 6.4				

[a] This table includes only isolates that are epidemiologically unrelated.
[b] The number of total unique spacers does not equal the sum of unique spacers in each serovar, because a unique spacer was sometimes present in more than one serovar.

equal epidemiologic concordance for *Salmonella* serovar Saintpaul, but lower epidemiologic concordance for *Salmonella* serovars Heidelberg and Muenchen (Table 7).

DISCUSSION

There are several important criteria to follow when selecting genetic markers to use in an MLST scheme. First, the selected genetic markers should exhibit adequate sequence variations to provide separation of unrelated strains (33). Second, genetic markers that provide epidemiologically meaningful information should be selected so that the MLST scheme can exhibit high epidemiologic concordance. Last but not least, genetic markers should be present in the genome within all strains of the species of interest. Previous studies demonstrated that MLST schemes based on *Salmonella enterica* housekeeping genes showed poor discriminatory power compared to PFGE (14, 24, 46). Inclusion of virulence genes into one published MLST scheme for subtyping *Salmonella* serovar Typhimurium increased discriminatory power to 0.98, which was comparable to that of PFGE (0.96) (15). Similarly, virulence genes provided epidemiologically meaningful separation and clustering of strains of *Listeria monocytogenes* (10). Besides virulence genes, CRISPRs were selected as markers in the current MLST scheme because they were found to be one of the fastest evolving genetic elements in bacterial genomes (40).

In the present study, cluster analyses based on the two virulence genes and two CRISPRs accurately grouped isolates according to their specific serovars, except for *Salmonella* serovar Typhimurium and *Salmonella* serovar I 4,[5],12:i:−, which were clustered together. As *Salmonella* serovar I 4,[5],12:i:− is a monophasic variant of *Salmonella* serovar Typhimurium (12), this was not unexpected. Virulence genes were previously found to provide accurate identification of different serovars of *Salmonella* in other studies as well (38, 44, 45).

Addition of CRISPRs significantly increased discriminatory power (Fig. 1) compared to previously published MLST schemes, and the identification of individual outbreak strains/clones was achieved. For example, one MLST scheme based on three housekeeping genes (*manB*, *pduF*, and *glnA*) and one virulence gene (*spaM*) identified one sequence type among 85 *Salmonella* serovar Typhimurium isolates and discriminatory power for the MLST scheme was 0 (14). Another MLST scheme targeted seven housekeeping genes, *aroC*, *dnaN*, *hemD*, *hisD*, *purE*, *sucA*, and *thrA*, and identified 12 sequence types among a total of 81 *Salmonella* serovar Newport isolates, which also resulted in poor discriminatory power (0.61) (24). One MLST study based on virulence genes (*hilA*, *pefB*, and *fimH*), 16S rRNA gene, and housekeeping genes showed high discriminatory power (0.98) for subtyping *Salmonella* serovar Typhimurium (15); however, its capacity to discriminate strains from more-clonal serovars, such as *Salmonella* serovar Enteritidis, was not tested. In conclusion, the MLST scheme described in the present study has superior discriminatory power compared to previously published MLST schemes for subtyping the major serovars of *Salmonella*, especially for the highly clonal *Salmonella* serovar Enteritidis.

As mentioned previously, the isolates selected for this study were biased toward those that had poor epidemiologic concordance by PFGE; therefore, future studies comparing MLST and PFGE need to be performed using a nonbiased strain

TABLE 7. Comparison of epidemiologic concordance between PFGE and MLST for the selected strains analyzed in the present study[a]

Subtyping method	Epidemiologic concordance between the two methods for the following *Salmonella* serovar:							
	Enteritidis	Typhimurium	Newport	Heidelberg	I 4,[5],12:i:−	Saintpaul	Montevideo	Muenchen
MLST	0.94	1.00	1.00	0.88	1.00	1.00	1.00	0.39
PFGE[b]	0.91	0.91	0.93	1.00	1.00	1.00	0.87	0.92

[a] Values for epidemiologic concordance were calculated based on isolates with cluster codes identified by PulseNet.
[b] The above values for epidemiologic concordance are biased against PFGE, because in some cases outbreaks that contained strains with variations in PFGE patterns (had poor epidemiologic concordance by PFGE) were deliberately selected in the present study.

collection. Generally speaking, the current MLST scheme showed high epidemiologic concordance for subtyping the major serovars of *Salmonella*, except for *Salmonella* serovar Muenchen ($E = 0.39$) (Table 7). All *Salmonella* serovar Muenchen isolates had different sequence types, except isolates SMcn13 and SMcn15 from the outbreak linked to orange juice (Table 1). Interestingly, the allelic types of *fimH* and *sseL* were the same for all the *Salmonella* serovar Muenchen isolates, except for isolate SMcn12 (Fig. 1a), which means that CRISPR1 and CRISPR2 provided almost all of the discriminatory power in the case of *Salmonella* serovar Muenchen isolates (Fig. 1b and c). One possible explanation for this unexpected diversity may be that CRISPRs are evolving too fast for investigations of *Salmonella* serovar Muenchen outbreaks, either because the specific niche where *Salmonella* serovar Muenchen resides harbors a large number of different phages and/or because phage pools of *Salmonella* serovar Muenchen are very dynamic. Dramatic differences have been observed in the rate of spacer acquisition between different eubacteria. In *Streptococcus thermophilus*, CRISPRs are very active, and new spacer acquisition appears to be the primary mechanism by which this species evolves phage resistance (11); however, the rate of new spacer acquisition in other bacteria, such as *Escherichia coli*, appears to be much slower (47). Alternatively, CRISPRs may have detected epidemiologically relevant differences in *Salmonella* serovar Muenchen isolates that were not detected by PFGE or discovered in the investigations of outbreaks.

The current MLST scheme also separated *Salmonella* serovar Enteritidis isolates with common PFGE patterns (Table 1). The predominant XbaI PFGE patterns for *Salmonella* serovar Enteritidis in the PulseNet database are JEGX01.0004 and JEGX01.0005 making up 45% and 13% of the database, respectively (CDC, unpublished data). This lack of PFGE pattern diversity makes it difficult to differentiate potential outbreak-related isolates from sporadic isolates (17). The discriminatory power of PFGE has been increased by the combination of multiple restriction enzymes (50). However, whether the increased discrimination provided by additional restriction enzymes caused potential loss of epidemiologic concordance was not addressed in that study. The present MLST scheme allowed separation of the two predominant PFGE patterns of *Salmonella* serovar Enteritidis isolates (Table 1) and resulted in high epidemiologic concordance (Table 7). CRISPRs provided most of the discrimination among outbreak strains/clones (Fig. 1b and c). CRISPRs in *Salmonella* serovar Enteritidis are evolving due to plasmids and/or phages present in the environment (48). Fortunately, the rate of spacer insertion and deletion in CRISPRs is slow enough such that they do not appear to change during the time course of an outbreak (Table 1). CRISPRs may also reflect the specific phage and plasmid pool in the environment and hence contain ecologically and geographically meaningful information for bacteria (28, 48) that may be useful for tracking strains of *Salmonella* to their specific source (farm or food processing plant). In conclusion, the current MLST scheme effectively subtyped the two most common PFGE patterns of *Salmonella* serovar Enteritidis and thus could enhance cluster definition and outbreak investigation capabilities of public surveillance laboratories.

It has been previously suggested that CRISPRs are poor epidemiological markers in enterobacteria due to the low rate of spacer acquisition (47). However, that study analyzed only 16 complete *Salmonella* genomes for CRISPRs, and only four of them were from the same serovar as strains analyzed in the current study. Additionally, the authors included in their study only one isolate from *Salmonella* serovars Typhimurium, Enteritidis, Newport, and Heidelberg, so the true value of CRISPRs for epidemiologic investigations could not be adequately evaluated given their sampling limitations. Our study analyzed 26, 34, 15, and 20 isolates from *Salmonella* serovars Typhimurium, Enteritidis, Newport, and Heidelberg, respectively, and demonstrated that CRISPR sequences may provide the discriminatory power and epidemiologic concordance needed for epidemiologic investigations. We are currently testing this hypothesis further using larger numbers of isolates obtained from current and past *Salmonella* outbreaks. The previously observed discrepancy between CRISPR sequences and strain phylogeny (47) suggests that the MLST method reported here would not be useful for determining the long-term phylogeny of *Salmonella* isolates.

This MLST scheme has several other advantages that make it a potential subtyping method for routine surveillance of *Salmonella*. First, the primers in this MLST scheme were designed to have the same annealing temperature for all four markers so that it can be conveniently performed in large-scale epidemiologic investigations. Second, the number of the markers targeted was minimized to two virulence genes and two CRISPRs so that time and expense can be saved during routine typing of *Salmonella* strains (33). Third, all four markers, *fimH*, *sseL*, CRISPR1, and CRISPR2, are present in the major serovars of *Salmonella* and also in all published genomes of *Salmonella* serovars, so the current MSLT scheme is widely applicable. Although this MLST scheme shows great promise, future research is needed to further validate it for epidemiologic purposes and to compare it to more commonly used molecular subtyping tools for *Salmonella*, including PFGE and MLVA. One advantage our method has over MLVA though is that it appears to be universally applicable to the most clinically relevant *Salmonella* serovars, where MLVA protocols for only a limited number of serovars have been described so far.

In conclusion, this study suggests that an MLST protocol that includes CRISPR1 and CRISPR2 may be a useful subtyping method for tracking the farm-to-fork spread of the most prevalent serovars of *Salmonella* during investigations of outbreaks.

ACKNOWLEDGMENTS

We thank Bindhu Verghese for technical guidance throughout the study, especially for the idea of combining CRISPRs into one allele for the cluster analysis. We also acknowledge the Penn State Genomics Core Facility-University Park, PA, for DNA sequencing and Eija Hytia-Trees at the CDC for assistance in selecting the strain collection.

This study was supported by a U.S. Department of Agriculture Special Milk Safety grant to the Pennsylvania State University (contract 2009-34163-20132).

REFERENCES

1. **Barrangou, R., et al.** 2007. CRISPR provides acquired resistance against viruses in prokaryotes. Science **315:**1709–1712.
2. **Beranek, A., et al.** 2009. Multiple-locus variable-number tandem repeat analysis for subtyping of *Salmonella enterica* subsp. *enterica* serovar Enteritidis. Int. J. Med. Microbiol. **299:**43–51.

3. **Brouns, S. J. J., et al.** 2008. Small CRISPR RNAs guide antiviral defense in prokaryotes. Science **321:**960–964.
4. **CDC.** 2006. *Salmonella* annual summary 2005. Division of Foodborne, Bacterial and Mycotic Disease, National Center for Zoonotic, Vector-Borne and Enteric Diseases, Coordinating Center for Infectious Diseases, Centers for Diseases Control and Prevention, U.S. Department of Health and Human Services, Atlanta, GA. http://www.cdc.gov/ncidod/dbmd/phlisdata/salmtab/2005/SalmonellaAnnualSummary2005.pdf.
5. **CDC.** 2008. *Salmonella* annual summary 2006. Division of Foodborne, Bacterial and Mycotic Disease, National Center for Zoonotic, Vector-Borne and Enteric Diseases, Coordinating Center for Infectious Diseases, Centers for Diseases Control and Prevention, U.S. Department of Health and Human Services, Atlanta, GA. http://www.cdc.gov/ncidod/dbmd/phlisdata/salmtab/2006/SalmonellaAnnualSummary2006.pdf.
6. **CDC.** 2008. Outbreak of *Salmonella* serotype Saintpaul infections associated with multiple raw produce items—United States, 2008. MMWR Morb. Mortal. Wkly. Rep. **57:**929–934.
7. **CDC.** 2 December 2010, final posting date. Investigation update: multistate outbreak of human *Salmonella* Enteritidis infections associated with shell eggs. Centers for Diseases Control and Prevention, Atlanta, GA. http://www.cdc.gov/salmonella/enteritidis/.
8. **CDC.** 4 May 2010, final posting date. Investigation update: multistate outbreak of human *Salmonella* Montevideo infections. Centers for Diseases Control and Prevention, Atlanta, GA. http://www.cdc.gov/salmonella/montevideo/index.html.
9. **CDC.** 2010. Preliminary FoodNet data on the incidence of infection with pathogens transmitted commonly through food—10 states, 2009. MMWR Morb. Mortal. Wkly. Rep. **59:**418–422.
10. **Chen, Y., W. Zhang, and S. J. Knabel.** 2007. Multi-virulence-locus sequence typing identifies single nucleotide polymorphisms which differentiate epidemic clones and outbreak strains of *Listeria monocytogenes*. J. Clin. Microbiol. **45:**835–846.
11. **Deveau, H., et al.** 2008. Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. J. Bacteriol. **190:**1390–1400.
12. **Echeita, M. A., S. Herrera, and M. A. Usera.** 2001. Atypical, *fljB*-negative *Salmonella enterica* subsp. *enterica* strain of serovar 4,5,12:i:− appears to be a monophasic variant of serovar Typhimurium. J. Clin. Microbiol. **39:**2981–2983.
13. **Endo, T., K. Ikeo, and T. Gojobori.** 1996. Large-scale search for genes on which positive selection may operate. Mol. Biol. Evol. **13:**685–690.
14. **Fakhr, M. K., L. K. Nolan, and C. M. Logue.** 2005. Multilocus sequence typing lacks the discriminatory ability of pulsed-field gel electrophoresis for typing *Salmonella enterica* serovar Typhimurium. J. Clin. Microbiol. **43:**2215–2219.
15. **Foley, S. L., et al.** 2006. Comparison of subtyping methods for differentiating *Salmonella enterica* serovar Typhimurium isolates obtained from food animal sources. J. Clin. Microbiol. **44:**3569–3577.
16. **Foley, S. L., S. Zhao, and R. D. Walker.** 2007. Comparison of molecular typing methods for the differentiation of *Salmonella* foodborne pathogens. Foodborne Pathog. Dis. **4:**253–276.
17. **Gerner-Smidt, P., et al.** 2006. PulseNet USA: a five-year update. Foodborne Pathog. Dis. **3:**9–19.
18. **Giammanco, G. M., et al.** 2007. Evaluation of a modified single-enzyme amplified fragment length polymorphism (SE-AFLP) technique for subtyping *Salmonella enterica* serotype Enteritidis. Res. Microbiol. **158:**10–17.
19. **Grimont, P. A. D., and F. Weill.** 2007. Antigenic formulae of the *Salmonella* serovars, 9th ed. WHO Collaborating Centre for Reference and Research on *Salmonella*. Institut Pasteur, Paris, France.
20. **Grissa, I., G. Vergnaud, and C. Pourcel.** 2008. CRISPRcompar: a website to compare clustered regularly interspaced short palindromic repeats. Nucleic Acids Res. **36:**W145–W148.
21. **Groenen, P. M., A. E. Bunschoten, D. Soolingen, and J. D. Errtbden.** 1993. Nature of DNA polymorphism in the direct repeat cluster of *Mycobacterium tuberculosis*: application for strain differentiation by a novel typing method. Mol. Microbiol. **10:**1057–1065.
22. **Hale, C. R., et al.** 2009. RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. Cell **139:**945–956.
23. **Hanning, I. B., J. D. Nutt, and S. C. Ricke.** 2009. Salmonellosis outbreaks in the United States due to fresh produce: sources and potential intervention measures. Foodborne Pathog. Dis. **6:**635–648.
24. **Harbottle, H., D. G. White, P. F. McDermott, R. D. Walker, and S. Zhao.** 2006. Comparison of multilocus sequence typing, pulsed-field gel electrophoresis, and antimicrobial susceptibility typing for characterization of *Salmonella enterica* serotype Newport isolates. J. Clin. Microbiol. **44:**2449–2457.
25. **Hoe, N., et al.** 1999. Rapid molecular genetic subtyping of serotype M1 group A *Streptococcus* strains. Emerg. Infect. Dis. **5:**254–263.
26. **Horvath, P., and R. Barrangou.** 2010. CRISPR/Cas, the immune system of bacteria and archaea. Science **327:**167–170.
27. **Hunter, P. R., and M. A. Gaston.** 1988. Numerical index of the discriminatory ability of typing systems: an application of Simpson's index of diversity. J. Clin. Microbiol. **26:**2465–2466.
28. **Kunin, V., et al.** 2008. A bacterial metapopulation adapts locally to phage predation despite global dispersal. Genome Res. **18:**293–297.
29. **Lan, R., P. R. Reeves, and S. Octavia.** 2009. Population structure, origins and evolution of major *Salmonella enterica* clones. Infect. Genet. Evol. **9:**996–1005.
30. **Lindstedt, B., et al.** 2007. Harmonization of the multiple-locus variable-number tandem repeat analysis method between Denmark and Norway for typing *Salmonella* Typhimurium isolates and closer examination of the VNTR loci. J. Appl. Microbiol. **102:**728–735.
31. **Lindstedt, B., E. Heir, E. Gjernes, and G. Kapperud.** 2003. DNA fingerprinting of *Salmonella enterica* subsp. *enterica* serovar typhimurium with emphasis on phage type DT104 based on variable number of tandem repeat loci. J. Clin. Microbiol. **41:**1469–1479.
32. **Lindstedt, B., E. Heir, T. Vardund, and G. Kapperud.** 2000. Fluorescent amplified-fragment length polymorphism genotyping of *Salmonella enterica* subsp. *enterica* serovars and comparison with pulsed-field gel electrophoresis typing. J. Clin. Microbiol. **38:**1623–1627.
33. **Maiden, M. C. J.** 2006. Multilocus sequence typing of bacteria. Annu. Rev. Microbiol. **60:**561–588.
34. **Pourcel, C., G. Salvignol, and G. Vergnaud.** 2005. CRISPR elements in *Yersinia pestis* acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. Microbiology **151:**653–663.
35. **Ribot, E. M., et al.** 2006. Standardization of pulsed-field gel electrophoresis protocols for the subtyping of *Escherichia coli* O157:H7, *Salmonella*, and *Shigella* for PulseNet. Foodborne Pathog. Dis. **3:**59–67.
36. **Ross, I. L., and M. W. Heuzenroeder.** 2005. Use of AFLP and PFGE to discriminate between *Salmonella enterica* serovar Typhimurium DT126 isolates from separate food-related outbreaks in Australia. Epidemiol. Infect. **133:**635–644.
37. **Ross, I. L., and M. W. Heuzenroeder.** 2009. A comparison of two PCR-based typing methods with pulsed-field gel electrophoresis in *Salmonella enterica* serovar Enteritidis. Int. J. Med. Microbiol. **299:**410–420.
38. **Scaria, J., et al.** 2008. Microarray for molecular typing of *Salmonella enterica* serovars. Mol. Cell. Probes **22:**238–243.
39. **Schouls, L. M., et al.** 2003. Comparative genotyping of *Campylobacter jejuni* by amplified fragment length polymorphism, multilocus sequence typing, and short repeat sequencing: strain diversity, host range, and recombination. J. Clin. Microbiol. **41:**15–26.
40. **Sorek, R., V. Kunin, and P. Hugenholtz.** 2008. CRISPR—a widespread system that provides acquired resistance against phage in bacteria and archaea. Nat. Rev. Microbiol. **6:**181–186.
41. **Struelens, M. J.** 1996. Consensus guidelines for appropriate use and evaluation of microbial epidemiologic typing systems. Clin. Microbiol. Infect. **2:**2–11.
42. **Tamada, Y., et al.** 2001. Molecular typing and epidemiological study of *Salmonella enterica* serotype Typhimurium isolates from cattle by fluorescent amplified-fragment length polymorphism fingerprinting and pulsed-field gel electrophoresis. J. Clin. Microbiol. **39:**1057–1066.
43. **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.
44. **Tankouo-Sandjong, B., et al.** 2007. MLST-v, multilocus sequence typing based on virulence genes, for molecular typing of *Salmonella enterica* subsp. *enterica* serovars. J. Microbiol. Methods **69:**23–36.
45. **Tankouo-Sandjong, B., et al.** 2008. Development of an oligonucleotide microarray method for *Salmonella* serotyping. Microb. Biotechnol. **1:**513–522.
46. **Torpdahl, M., M. N. Skov, D. Sandvang, and D. L. Baggesen.** 2005. Genotypic characterization of *Salmonella* by multilocus sequence typing, pulsed-field gel electrophoresis and amplified fragment length polymorphism. J. Microbiol. Methods **63:**173–184.
47. **Touchon, M., and P. C. E. Rocha.** 2010. The small, slow and specialized CRISPR and anti-CRISPR of *Escherichia* and *Salmonella*. PLoS One **5:**e11126.
48. **Vale, P. F., and T. J. Little.** 2010. CRISPR-mediated phage resistance and the ghost of coevolution past. Proc. R. Soc. Lond. B Biol. Sci. **277:**2097–2103.
49. **Voetsch, A., et al.** 2004. FoodNet estimate of the burden of illness caused by nontyphoidal *Salmonella* infections in the United States. Clin. Infect. Dis. **38:**127–134.
50. **Zheng, J., C. E. Keys, S. Zhao, J. Meng, and E. W. Brown.** 2007. Enhanced subtyping scheme for *Salmonella* Enteritidis. Emerging Infect. Dis. **13:**1932–1935.