

Diverse Genetic Markers Concordantly Identify Bovine Origin *Escherichia coli* O157 Genotypes Underrepresented in Human Disease^{∇†}

Joshua Whitworth,¹ Yubei Zhang,¹ James Bono,² Eve Pleydell,³ Nigel French,³ and Thomas Besser^{1*}

Department of Veterinary Microbiology and Pathology, Washington State University College of Veterinary Medicine, Pullman, Washington 99164¹; U.S. Meat Animal Research Center, Agricultural Research Service, U.S. Department of Agriculture, Clay Center, Nebraska 68933-0166²; and Institute of Veterinary, Animal, and Biomedical Sciences, Massey University, Palmerston North, New Zealand³

Received 24 July 2009/Accepted 19 October 2009

Genetic markers previously reported to occur at significantly different frequencies in isolates of *Escherichia coli* O157:H7 obtained from cattle and from clinically affected humans concordantly delineate at least five genetic groups. Isolates in three of these groups consistently carry one or more markers rarely found among clinical isolates.

Escherichia coli serotype O157:H7 is an important zoonotic pathogen that may cause diarrhea, bloody diarrhea, and hemolytic-uremic syndrome (3, 10, 14). *E. coli* O157 is transmitted to humans by direct contact with the animal reservoir (which includes cattle and other ruminant animals) or indirectly by ingestion of contaminated food or water (3, 10). Genetic analyses of bovine isolates of *E. coli* O157 from diverse geographic origins have provided evidence for the global dissemination of genotypes and also for significant regional differences in the relative prevalence of some genotypes (5, 8, 18, 19).

Several research groups have identified genetic markers that occur at different relative frequencies among *E. coli* O157 isolates from human clinical cases and from cattle. One group initially used octamer-based genomic scanning to identify two lineages of U.S. origin *E. coli* O157 (7), of which lineage I was composed mostly (36/44) of clinical isolates and lineage II was composed mostly (25/32) of cattle isolates. Subsequently, a simpler multiplex PCR-based assay (Table 1), the lineage-specific polymorphism assay (LSPA), was developed to identify these lineages (19). Six LSPA loci with alleles characteristic of lineage I, lineage II, or neither lineage I nor lineage II are, respectively, classified with the digit 1, 2, or 3, and these digits are concatenated to an LSPA code: 111111 indicates lineage I, and 211111 indicates a genetically intermediate group termed lineage I/II (20), whereas all other genotype variations are considered to belong to lineage II. More recently, a typing assay based on Shiga toxin-encoding bacteriophage insertion (SBI) sites grouped 91 of 92 clinical *E. coli* O157 isolates from the northwestern United States into three clusters, of which

clusters 1 and 3 predominated (>90%) (16). SBI consists of six PCRs (Table 1) that amplify the Stx toxin genes and the insertion site junctions of the Stx1- and Stx2-encoding bacteriophages of *E. coli* O157. In a subsequent study, the predominance (92.6%) of clusters 1 and 3 was confirmed in 190 additional human (clinical) isolates (1). In contrast, many (48.8%) *E. coli* O157 isolates from cattle in the northwestern United States and western Canada demonstrated SBI patterns rarely found among human (clinical) isolates (1).

Additional individual markers reported to occur at differing frequencies among clinical and reservoir isolates include the presence or absence of *stx2-Q* junction alleles (e.g., Q933 and Q21 alleles in 90% and 15.2% of 66 human isolates versus 44% and 64.8% of 91 bovine isolates, respectively) (9) and the nonsynonymous single nucleotide polymorphism (SNP) 255T→A in *tir*, a key virulence gene of *E. coli* O157:H7 (<1% of 108 human isolates versus 44% of 77 bovine isolates had the A allele) (2).

The goal of this study was to evaluate the concordance of these various markers reported to occur at different frequencies among isolates from asymptomatic cattle and from human patients. A convenience set of 145 *E. coli* O157 isolates obtained from cattle, aggregated from two isolate sets chosen to maximize the diversity of geographic and temporal origins within our isolate bank and whose provenance and SBI types were described previously, was used for this study (1, 5, 18). Briefly, these isolates were non-sorbitol-fermenting, beta-glucuronidase-negative *E. coli* O157 isolates from cattle on 130 different premises in five countries and 14 U.S. states, isolated in 12 different years ranging from 1991 through 2004. The isolates from outside North America included isolates from Australia ($n = 7$, obtained in 1993 to 2003), Japan ($n = 17$, obtained in 1996 to 1997), and Scotland ($n = 11$, obtained in 1999). LSPA was applied to this set by using previously described primer sequences (19), although capillary rather than gel electrophoresis was used (Table 1; DNA analyzer 3730, LIZ 600 size standard; Applied Biosystems, Foster City, CA). Data were analyzed with GeneMarker software (SoftGenetics,

* Corresponding author. Mailing address: Department of Veterinary Microbiology and Pathology, College of Veterinary Medicine, Washington State University, P.O. Box 647040, Pullman, WA 99164-7040. Phone: (509) 335-6075. Fax: (509) 335-8529. E-mail: tbesser@vetmed.wsu.edu.

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

∇ Published ahead of print on 30 October 2009.

TABLE 1. Oligonucleotides used in this study

Test	Target	Type	Sequence (5'→3') ^a	Reference(s)			
LSPA	<i>Fold-sfmA</i> <i>Z5935</i> <i>yhcG</i> <i>rbsB</i> <i>rtcB</i> <i>arp-iclR</i>	Primer	VIC-TACGTAGGTCAAGGG	18, 20			
		Primer	CCAGATTTACAACGCC				
		Primer	FAM-GTGTCCCCGGTATTTG				
		Primer	CTCACTGGCGTAACTT				
		Primer	VIC-CTCTGCAAAAACTTACGCC				
		Primer	CAGGTGGTTGATCAGCG				
		Primer	FAM-AGTTAATGTTCTTGCCAGCC				
		Primer	ATTCACCGCTTTTTCGCC				
		Primer	VIC-GCGCCAGATCGATAAAGTAAG				
		Primer	GCCGTTGTAAACGTGATAAAG				
		Primer	FAM-GCTCAATCTCATAATGCAGCC				
		Primer	CACGTATTACCGATGACCG				
		SBI	<i>stx1</i>		Primer	CGCTTTGCTGATTTTTCACA	16
					Primer	GTAACATCGCTCTTGCCACA	
<i>stx2</i>	Primer		GTTCCGGAATGCAAATCAGT				
	Primer		CGGCGTCATCGTATACACAG				
L <i>yehV</i> -phage	Primer		CACCGGAAGGACAATTCATC				
R <i>yehV</i> -phage	Primer		AACAGATGTGTGGTGAAGTGTCTG				
L <i>wrbA</i> -phage	Primer		AAGTGGCGTTGCTTTGTGAT				
R <i>wrbA</i> -phage	Primer		GATGCACAATAGGCACTACGC				
Primer	CCGACCTTTGTACGGATGTAA						
Primer	CGAATCGCTACGGAATAGAGA						
Primer	AGGAAGGTACGCATTTGACC						
Primer	ATCGTTCGCAAGAATCACAA						
Q933 Q21	Q933F <i>stx2a</i> R Q21F <i>stx2a</i> R	Primer	CGGAGGGGATTGTTGAAGGC	9			
		Primer	CCGAAGAAAAACCCAGTAACAG				
		Primer	GAAATCCTCAATGCCTCGTTG				
		Primer	CCGAAGAAAAACCCAGTAACAG				
Tir	<i>tirF</i> <i>tirR</i> <i>tirA</i> <i>tirT</i>	Primer	TGGCGGCGTCTGAGATAAC	2			
		Primer	GAGTATCGAGCGGACCATGATC				
		Probe	VIC-ACTGAATGATGGATTTG-MGBNFQ				
		Probe	FAM-CTGAATGAAGGATTTG-MGBNFQ				
		Probe	FAM-CTGAATGAAGGATTTG-MGBNFQ				

^a VIC, proprietary reporter dye, Applied Biosystems; FAM, 6-carboxyfluorescein; MGBNFQ, molecular-groove-binding nonfluorescent quencher.

LLC, State College, PA). *Q-stx2* alleles Q933 and Q21 were detected by PCR, and the *tir* polymorphism was detected by real-time PCR as described previously (2, 9).

Comparison of typing results produced by the LSPA, SBI, *Q-stx2*, and *tir* methods showed considerable overall agreement. Cross-classification of the LSPA and SBI results (Table 2) showed particularly strong agreement in assignment to the two human disease-associated genotypes (LSPA 111111 and 211111; SBI 1 and 3; chi square = 268, 20 df, $P < 0.001$; Cramer's V statistic = 0.681). *Q-stx2* typing identified the Q933 allele in 117 isolates, including 59 of 60 LSPA/SBI human disease-associated genotypes. The Q21 allele was detected in 67 isolates but was not strongly associated with either human disease or cattle-associated genotypes overall (data not shown). The *tir* nucleotide 255A allele was detected in 39 isolates, only 1 of which had an LSPA/SBI human disease-associated genotype.

While these cross-comparisons supported a significant degree of concordance between the results of the various typing systems, the data analysis was complicated by the differing numbers of genotypes determined by the different systems, and in particular by the classification by LSPA and SBI of numerous isolates into a number of sparsely populated genotypes (Table 2). More generally, it seemed likely that the best classification of the isolates would result from a consideration of all of the data generated. Therefore, we used Markov chain

Monte Carlo (MCMC) model-based clustering, implemented in the *structure* software package, version 2.2 (6), to investigate the population structure using as input data the 15 locus-specific test results (i.e., the six loci each from the LSPA and

TABLE 2. Cross-tabulation of genotypes identified by the SBI and LSPA methods among 145 isolates of bovine *E. coli* O157:H7 isolates of diverse temporal and geographic origins

LSPA type	No. of isolates typed by SBI as:					Total
	1	3	5	6	Other ^a	
111111	0	44	0	2	10	56
211111	16	0	2	1	8	27
213111	0	0	1	13	2	16
222213	1	0	5	1	0	7
222222	0	0	2	5	0	7
221213	2	0	4	0	0	6
Other ^b	2	1	20	2	1	26
Total	21	45	34	24	21	145

^a SBI also identified five isolates of genotype 7; four isolates of genotype 10; three isolates of genotype 16; two isolates each of genotypes 11, 14, and 15; and single isolates of genotypes 4, 12, and 13 (1).

^b LSPA also identified three isolates each of genotypes 222113 and 222313; two isolates each of genotypes 212111, 222212, 223213, 231111, 231233, and 232233; and single isolates of genotypes 111211, 212113, 222223, 223212, 223313, and 232233 (20).

SBI genotyping panels together with the Q933, Q21, and *tir* loci (see Table S1 in the supplemental material). The model assumes *K* populations, each of which is characterized by allele frequencies at multiple unlinked or weakly linked loci. Within each population, the loci are assumed to be at linkage equilibrium. It was not possible to test the validity of these assumptions for the isolate set modeled here, and it is likely that at least some degree of linkage disequilibrium is present within *E. coli* O157:H7 populations (12). We utilized this model both to determine the most likely number of populations (*K*) within the isolate set and to assign individual isolates to the best-fitting population(s). *K* = 1 would imply a lack of genetic substructure within the isolate set, while any *K* of >1 would assume the presence of the corresponding number of subgroups with distinct sets of allele frequencies. Initial assignments of group membership for each isolate were based on the location (North America, Scotland, Japan, or Australia) of the cattle from which the *E. coli* O157 isolates were obtained, due to the potential for genetic divergence of geographically separated populations.

K values of 1 to 10 were initially evaluated with 10 model runs each, with each run consisting of a 20,000-step burn-in followed by a 50,000-step parameter estimation. Comparison of the estimated logarithmic posterior probabilities [$\ln P(X|K)$, where *X* is the data] of these runs revealed that *K* values of <4 or >7 were highly unlikely. Additional runs (25 runs, each consisting of 100,000 steps for burn-in, followed by 100,000 steps for parameter estimation), were then performed in order to model each *K* value from 4 through 7. The results of these models demonstrated nearly equal maximum relative posterior probabilities for *K* = 5 and *K* = 6.

We selected *K* = 5 models for assigning isolates to specific clusters, based on (i) the parsimony principle (*K* = 5 being a less complex population structure than *K* = 6), (ii) the precision of the posterior probabilities (*K* = 5 models had consistently lower variances than *K* = 6 models), (iii) the lack of sensitivity of the model-derived posterior probabilities to the prior population assignments used to initialize the model (posterior probabilities of models initialized or not initialized with each isolate's country of origin increasingly diverged in values as *K* increased from 6), and (iv) the admixture determinations for individual isolates (as *K* increased from 6, an increasing proportion of the study isolates shared characteristics of two or more clusters). Cluster assignments from six independent, randomly selected *K* = 5 model runs were compared for concordance: using a criterion of a 0.5 or higher probability to assign isolates to their best-fit clusters, all cluster assignments from the six selected runs were perfectly concordant, with 140 to 142 isolates assigned to specific clusters, leaving only 3 to 5 isolates (depending on the run) with no cluster assignable at a 0.5 or higher probability (see Fig. S1 in the supplemental material). However, it is possible that the uncertainty of these ancestry assignments was underestimated or that the assignments were biased as a result of possible violation of the assumptions of linkage equilibrium within populations (6).

The concordant assignments of 142 isolates to five genetic clusters (designated A to E) were then used as the basis for individual evaluation of the different genetic typing systems by comparing each genotyping test or system for agreement with the model-derived cluster assignments. These comparisons re-

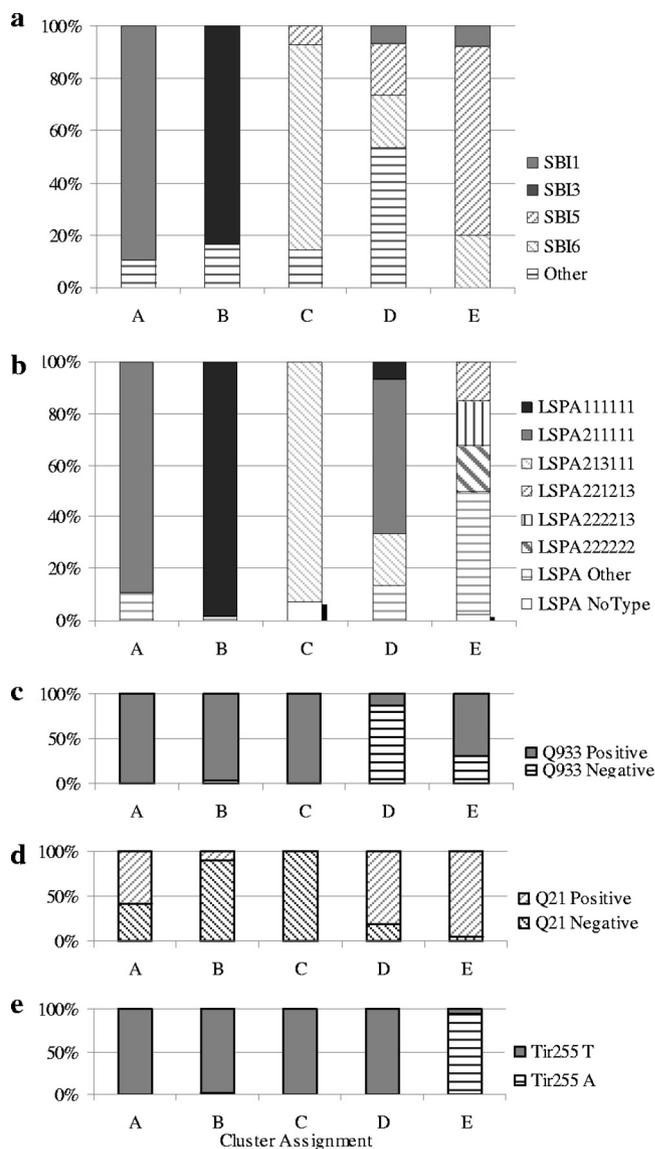


FIG. 1. MCMC model-based genetic cluster assignments (A to E) and their association with (a) SBI typing, (b) LSPA typing, (c) Q933 typing, (d) Q21 typing, and (e) *tir* typing. Clusters A to E included 19, 54, 14, 15, and 40 isolates, respectively.

vealed associations between genetic markers typical of human infection (for example, SBI type 1 and LSPA type 211111 in cluster A and SBI type 3 and LSPA type 111111 in cluster B), whereas isolates in clusters C to E each contained one or more markers rarely found in clinical isolates (Fig. 1). All markers/marker systems were strongly nonindependently distributed among the model-derived clusters ($\chi^2 = 84$ to 338; 4 to 16 df, Cramer's *V* = 0.662 to 0.937; *P* < 0.001 for each system). Not surprisingly, some isolates were assigned to clusters C to E by the model based on the complete data set despite carrying one or more markers typical of clinical isolates. For example, LSPA type 211111 was frequent among isolates assigned to both clusters A (17 of 19) and D (9 of 15), suggesting that this LSPA genotype may be polyphyletic. Clusters A and B cumulatively contained 73 of the 142 classified isolates (51%). We previ-

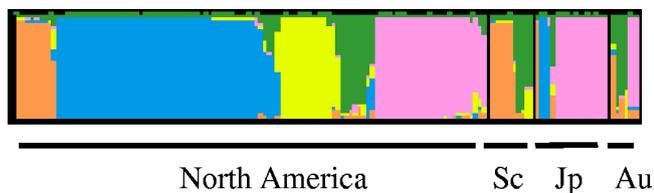


FIG. 2. MCMC model-based genetic cluster assignments and their association with regions of origin. Each isolate is depicted as a single vertical bar colored to represent its genetic cluster admixture (orange, cluster A; blue, cluster B; yellow, cluster C; green, cluster D; pink, cluster E). The isolates are sorted by cluster and location of isolation. The locations are North America ($n = 108$), Scotland (Sc; $n = 10$), Japan (Jp; $n = 17$), and Australia (Au; $n = 7$). This image was generated by using the DISTRUCT 1.1 software (15).

ously reported that the proportions of isolates with SBI genotypes typical of clinical isolates in different countries was weakly correlated to the respective national incidences of *E. coli* O157:H7-associated hemolytic-uremic syndrome (18). The *structure* version 2.2-derived cluster assignments reported here also differed by isolate provenance (Fig. 2; $\chi^2 = 30.0$, 4 df, $P < 0.001$; Cramer's $V = 0.262$). While the number of international source isolates examined here is clearly insufficient to support strong inferences, the data indicate the possibilities of (i) the unique occurrence of cluster C in North America, (ii) a relatively high frequency of cluster A and a low frequency of cluster E in Scotland, and (iii) a relatively low frequency of cluster A in Japan and Australia. As the genetic markers of cluster A have been associated with increased virulence (11), further research on the association of the distribution of *E. coli* O157:H7 genotypes and the national incidence and severity of *E. coli* O157:H7-associated disease may be merited.

Multiple-correspondence analysis (MCA) and hierarchical clustering were used in a second approach to explore the relationships between the isolates defined by the same set of genetic markers by using the methods of Murtagh (13). The application of MCA provided an opportunity to test whether the clusters identified by the MCMC models were supported by this very different analytical method. MCA identifies a lower-dimensional subspace that approximately represents the diversity within a multivariate data set. In initial MCA models using the full data set, uncommon LSPA and SBI types (specifically, those each comprising less than 5% of the isolate set) exhibited a strong tendency to cocluster, and therefore these unusual types were pooled to produce four LSPA categories. MCA of this reduced data set (SBI [1, 3, 5, 6, or other], LSPA [111111, 211111, 213111, or other], Q933 [positive or negative], Q21 [positive or negative], and Tir [255T or 255A]) identified four dimensions (factors) that cumulatively accounted for >80% of the variation within the data set and retained 69 to 93% of the quality of representation of each marker (13) (see Table S2 in the supplemental material). The coordinates of the projections of each marker onto these four dimensions were extracted from the model and hierarchically clustered by using minimum-variance methods, weighting each marker by its mass (marginal total) (13) (Fig. 3), resulting in five clusters very similar (and named accordingly) to those produced by the MCMC model illustrated in Fig. 1.

In summary, these results clearly demonstrate that the sev-

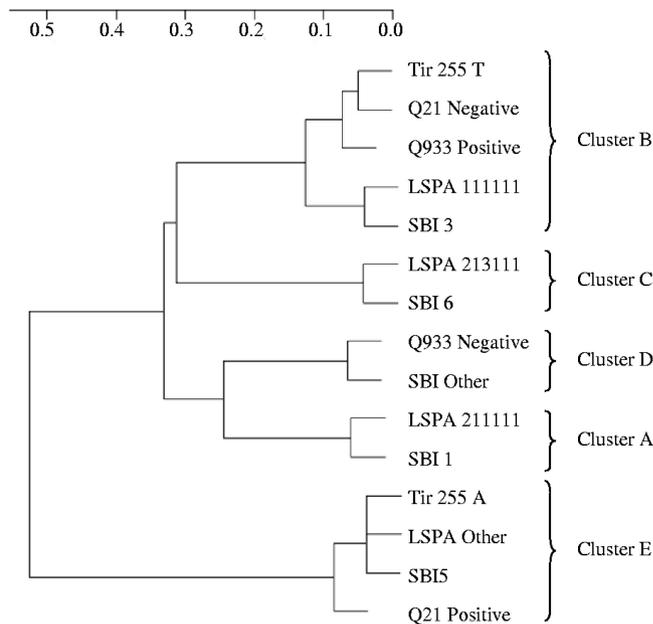


FIG. 3. MCA with hierarchical clustering of *E. coli* O157:H7 genotyping data, weighted to reflect the contributions of individual factors to the total inertia ($n = 138$ isolates with no missing data, Ward's minimum-variance method).

eral individual genetic tests or multiple test marker systems previously reported to occur at different frequencies among isolates from cattle and humans identify largely concordant genotypes of *E. coli* O157. The distribution of these markers among this international collection of isolates strongly indicated the existence of five (or more) genetic groups of *E. coli* O157, only two of which (clusters A and B) predominantly carry markers previously associated with clinical isolates. It is nearly certain that additional genetic groups or subgroups of *E. coli* O157 exist in nature, since delineation of these five groups is based on the sampling of only a tiny proportion of the genome: For example, in a recent study, 96 SNPs differentiated clinical *E. coli* O157 isolates into nine discrete clades (11). The Stx content and the relative frequencies of the two numerically predominant clades of clinical isolates identified in reference 11, clades 2 and 8, are consistent with those of clusters B and A, respectively, as described here.

The concordance of the multiple genetic markers, each with alleles differentially associated with human disease, supports the hypothesis of the existence of discrete genotypes of *E. coli* O157 that differ in their virulence for humans. This diversity is consistent with a source-sink ecological model characterized by broad genetic diversity in the reservoir (source) bovine populations that includes at least five genetic clusters, of which only two carry genetic markers typical of clinical isolates (17). In this ecological model, human infections represent a "sink" characterized by relatively short-duration infections unlikely to be persistently transmitted ($R_0 < 1.0$). The source-sink model implies that various *E. coli* O157 genotypes diverged in the bovine reservoir through genetic drift and/or through bovine fitness-based selection, during which some genotypes evolved into accidental human pathogens. Based on this model, we predict that the genomic DNA sequences of *E. coli* O157

genotypes largely restricted to the bovine reservoir will reveal more genetic diversity than is apparent from the clinical isolate sequences now available, and SNP data supporting this prediction have already appeared (4). Investigation of the presence and expression of virulence factors by diverse bovine *E. coli* O157 genotypes may be required to reveal the mechanism(s) underlying their differential association with human disease.

This work was funded in part by NIAID NIH contract N01-AI-30055 and by the Agricultural Animal Health Program, Washington State University College of Veterinary Medicine, Pullman.

REFERENCES

- Besser, T. E., N. Shaikh, N. J. Holt, P. I. Tarr, M. E. Konkel, P. Malik-Kale, C. W. Walsh, T. S. Whittam, and J. L. Bono. 2007. Greater diversity of Shiga toxin-encoding bacteriophage insertion sites among *Escherichia coli* O157:H7 isolates from cattle than in those from humans. *Appl. Environ. Microbiol.* **73**:671–679.
- Bono, J. L., J. E. Keen, M. L. Clawson, L. M. Durso, M. P. Heaton, and W. W. Laegreid. 2007. Association of *Escherichia coli* O157:H7 tir polymorphisms with human infection. *BMC Infect. Dis.* **7**:98–106.
- Caprioli, A., S. Morabito, H. Brugere, and E. Oswald. 2005. Enterohaemorrhagic *Escherichia coli*: Emerging issues on virulence and modes of transmission. *Vet. Res.* **36**:289–311.
- Clawson, M. L., J. E. Keen, T. P. Smith, L. M. Durso, T. G. McDanel, R. E. Mandrell, M. A. Davis, and J. L. Bono. 22 May 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. [Epub ahead of print.]
- Davis, M. A., D. D. Hancock, T. E. Besser, D. H. Rice, C. J. Hovde, R. Digiacomo, M. Samadpour, and D. R. Call. 2003. Correlation between geographic distance and genetic similarity in an international collection of bovine faecal *Escherichia coli* O157:H7 isolates. *Epidemiol. Infect.* **131**:923–930.
- Falush, D., M. Stephens, and J. K. Pritchard. 2003. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* **164**:1567–1587.
- Kim, J., J. Nietfeldt, 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.
- Kim, J., J. Nietfeldt, J. Ju, J. Wise, N. Fegan, P. Desmarchelier, and A. K. Benson. 2001. Ancestral divergence, genome diversification, and phylogeographic variation in subpopulations of sorbitol-negative, beta-glucuronidase-negative enterohemorrhagic *Escherichia coli* O157. *J. Bacteriol.* **183**:6885–6897.
- Lejeune, J. T., S. T. Abedon, K. Takemura, N. P. Christie, and S. Sreevatsan. 2004. Human *Escherichia coli* O157:H7 genetic marker in isolates of bovine origin. *Emerg. Infect. Dis.* **10**:1482–1485.
- Mainil, J. G., and G. Daube. 2005. Verotoxigenic *Escherichia coli* from animals, humans and foods: who's who? *J. Appl. Microbiol.* **98**:1332–1344.
- 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. U. S. A.* **105**:4868–4873.
- Milkman, R. 1997. Recombination and population structure in *Escherichia coli*. *Genetics* **146**:745–750.
- Murtagg, F. 2005. Correspondence analysis and data coding with Java and R. CRC Press, Boca Raton, FL.
- Rangel, J. M., P. H. Sparling, C. Crowe, P. M. Griffin, and D. L. Swardlow. 2005. Epidemiology of *Escherichia coli* O157:H7 outbreaks, United States, 1982–2002. *Emerg. Infect. Dis.* **11**:603–609.
- Rosenberg, N. A. 2004. DISTRUCT: a program for the graphical display of population structure. *Mol. Ecol. Notes* **4**:137–138.
- Shaikh, N., and P. I. Tarr. 2003. *Escherichia coli* O157:H7 Shiga toxin-encoding bacteriophages: integrations, excisions, truncations, and evolutionary implications. *J. Bacteriol.* **185**:3596–3605.
- Sokurenko, E. V., R. Gomulkiewicz, and D. E. Dykhuizen. 2006. Source-sink dynamics of virulence evolution. *Nat. Rev. Microbiol.* **4**:548–555.
- Whitworth, J. H., N. Fegan, J. Keller, K. S. Gobius, J. L. Bono, D. R. Call, D. D. Hancock, and T. E. Besser. 2008. International comparison of clinical, bovine, and environmental *Escherichia coli* O157 isolates on the basis of Shiga toxin-encoding bacteriophage insertion site genotypes. *Appl. Environ. Microbiol.* **74**:7447–7450.
- Yang, Z., J. Kovar, J. Kim, J. Nietfeldt, 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, beta-glucuronidase-negative *Escherichia coli* O157:H7 from bovine production environments and human clinical samples. *Appl. Environ. Microbiol.* **70**:6846–6854.
- Zhang, Y., C. Laing, M. Steele, K. Ziebell, R. Johnson, A. K. Benson, E. Taboada, and V. P. Gannon. 2007. Genome evolution in major *Escherichia coli* O157:H7 lineages. *BMC Genomics* **8**:121–137.