Multilocus Genotype Analysis of *Escherichia coli* O157 Isolates from Australia and the United States Provides Evidence of Geographic Divergence

Glen E. Mellor,^a Thomas E. Besser,^b Margaret A. Davis,^b Brittany Beavis,^b WooKyung Jung,^b Helen V. Smith,^c Amy V. Jennison,^c Christine J. Doyle,^c P. Scott Chandry,^d Kari S. Gobius,^d Narelle Fegan^d

CSIRO Animal, Food and Health Sciences, Archerfield BC, QLD, Australia^a; Washington State University, Department of Veterinary Microbiology and Pathology, Pullman, Washington, USA^b; Queensland Health Forensic and Scientific Services, Archerfield BC, QLD, Australia^c; CSIRO Animal, Food and Health Sciences, Werribee, VIC, Australia^d

Escherichia coli O157 is a food-borne pathogen whose major reservoir has been identified as cattle. Recent genetic information has indicated that populations of *E. coli* O157 from cattle and humans can differ genetically and that this variation may have an impact on their ability to cause severe human disease. In addition, there is emerging evidence that *E. coli* O157 strains from different geographical regions may also be genetically divergent. To investigate the extent of this variation, we used Shiga toxin bacteriophage insertion sites (SBI), lineage-specific polymorphisms (LSPA-6), multilocus variable-number tandem-repeat analysis (MLVA), and a *tir* 255T>A polymorphism to examine 606 isolates representing both Australian and U.S. cattle and human populations. Both uni- and multivariate analyses of these data show a strong association between the country of origin and multilocus genotypes ($P < 0.0001$). In addition, our results identify factors that may play a role in virulence that also differed in isolates from each country, including the carriage of *stx*₁ in the *argW* locus uniquely observed in Australian isolates and the much higher frequency of *stx*₂-positive (also referred to as *stx*_{2a}) strains in the U.S. isolates (4% of Australian isolates versus 72% of U.S. isolates). LSPA-6 lineages differed between the two continents, with the majority of Australian isolates belonging to lineage I/II (LI/II) (LI, 2%; LI/II, 85%; LII, 13%) and the majority of U.S. isolates belonging to LI (LI, 60%; LI/II, 16%; LII, 25%). The results of this study provide strong evidence of phylogeographic structuring of *E. coli* O157 populations, suggesting divergent evolution of enterohemorrhagic *E. coli* O157 in Australia and the United States.

Escherichia coli O157:H7 (O157) is a food-borne pathogen that primarily resides in the gastrointestinal tract of healthy ruminants, particularly cattle (1–3). Transmission to humans can occur through the consumption of contaminated foods and water, person-to-person contact, or contact with an animal reservoir (3). Infection can be asymptomatic and can progress to uncomplicated gastroenteritis or hemorrhagic colitis and may be followed by a severe sequela known as hemolytic uremic syndrome (HUS). In the United States during 2006, Shiga-toxigenic *E. coli* O157 (STEC O157) infected an estimated 96,534 people, resulting in 3,268 hospitalizations and 31 fatalities (4). The total burden of STEC O157 on the U.S. health care system has been estimated at \$405 million per annum, of which a significant portion was linked to HUS patients (5).

E. coli O157 is the primary cause of HUS in many regions and countries, including North America, Scotland, and Argentina (2, 6–8). Interestingly, though, the frequencies of HUS and incidences of infection differ substantially worldwide (9–13). In North America during 2010, the rate of infection in the population was estimated at 0.9 cases per 100,000 (11). While this represents a significant reduction from previous years, the incidence of infection in the United States remains 7-fold higher than the incidence in Australia (0.12 cases per 100,000) (9). Furthermore, the rate of infection in Scotland has been reported at 4.4 cases per 100,000, 36-fold higher than Australia (10). Although there is currently no direct information available on the frequency of infection in Argentina, the incidence of HUS in that country is among the highest in the world (12) and most cases of HUS in this region have been attributed to *E. coli* O157 (6, 12, 14).

The prevalence of *E. coli* O157 in Australian cattle feces (1.5%

to 13%) is similar to that reported in Scotland (7.5%), Argentina (4.1%), and the United States (<1% to >25%) (1, 2, 6, 15–17). The similar rates of cattle prevalence but different rates of human infection between countries may result from many factors, such as disease surveillance, food-handling practices, supply chains, and host susceptibility. In addition, genetic diversity of *E. coli* O157 is also gaining credibility as a potential determinant of virulence (18–21). To date, several studies have identified variations in the incidence and severity of disease associated with different *E. coli* O157 outbreak-associated strains (19, 22, 23). Manning et al. examined the genetic relationship of a large set of clinical isolates responsible for a variety of disease manifestations in patients from North America (19). Their findings demonstrate a significant association between the severity and duration of disease and genetically related clades (19). Further studies have also identified clinically significant genotypes using techniques to differentiate isolates from human and bovine origins (20, 21, 24–26). Bono et al. identified strains of *E. coli* O157 frequently isolated from cattle but rarely from human disease that carry alternative single nucleotide polymorphism (SNP) alleles within *tir*, the gene encoding

Received 10 May 2013 Accepted 12 June 2013

Published ahead of print 14 June 2013

Address correspondence to Narelle Fegan, Narelle.Fegan@csiro.au.

Supplemental material for this article may be found at <http://dx.doi.org/10.1128/AEM.01525-13>.

Copyright © 2013, American Society for Microbiology. All Rights Reserved.
doi:10.1128/AEM.01525-13

the translocated intimin receptor that enables the characteristic adherence of this organism to epithelial cells (21). Evidence for source attribution differences is also provided by epidemiological studies from North America and Japan which indicate the dominance of specific lineages in isolates derived from human or cattle reservoirs (20, 24, 25, 27).

The evolution of *E. coli* O157 occurred in a stepwise manner through the loss and acquisition of various traits from an ancestral *E. coli* O55:H7 (clonal complex A1) strain to the typical sorbitol-negative, β -glucuronidase-negative *E. coli* O157 strain of clonal complex A6 (28, 29). Divergence between ancestral *E. coli* O55:H7 and contemporary *E. coli* O157:H7 is thought to have occurred relatively recently, with modern estimates suggesting that divergence occurred anywhere between 400 and just over 10,000 years ago (30, 31). Despite this recent divergence, the A6 clonal complex and its descendants are the most common *E. coli* O157:H7 clones found in animals and humans along with being the most geographically dispersed (32, 33). Continuing analysis of the A6 clonal lineage has identified genetic differences which suggest further differentiation into subclones (29). There is currently limited information available from small data sets which indicate that genetic diversity among the A6 clonal complex of *E. coli* O157 may be related to geographic separation (27, 32, 34). It is possible that these differences may in turn correlate with the observed variations in the transmissibility of *E. coli* O157 between animal reservoirs and humans or in the virulence of *E. coli* O157 in humans.

In this study, we examined the genetic relatedness of large diverse collections of cattle and human *E. coli* O157 from two geographically separate regions, Australia and the United States, by interrogating isolates for Shiga toxin bacteriophage insertion sites (SBI), lineage-specific polymorphisms (LSPA-6), multilocus tandem repeats (using multilocus variable-number tandem-repeat analysis [MLVA]), and a *tir* 255A>T single nucleotide polymorphism. These techniques have previously proven useful in discriminating *E. coli* O157 genotypes of various levels of clinical significance and were selected for inclusion on this basis. We suggest that there is a geographically biased distribution of genetically diverse *E. coli* O157 isolates and that this geographic divergence may have an impact on disease potential.

MATERIALS AND METHODS

Strain selection. A total of 606 *E. coli* O157 isolates from Australian cattle ($n = 205$), Australian humans ($n = 79$), U.S. cattle ($n = 143$), and U.S. humans ($n = 179$), in addition to reference strains *E. coli* O157 Sakai and *E. coli* O55 (660-79 and 5905), were selected for inclusion in this study. Where possible, isolates were chosen to maximize geographical, temporal, and genetic diversity. Australian and U.S. cattle isolates were collected from 6 states across a 16-year period (1993 to 2009) and 14 states over a 12-year period (1992 to 2004), respectively. Both isolate sets were chosen to represent a large number of different pulsed-field gel electrophoresis (PFGE) profiles.

A total of 225 *E. coli* O157-related illnesses, representing sporadic and outbreak-related cases, were notified in Australia between 2001 and 2009 (9). In this study, 79 Australian human isolates were obtained from 4 states across a period of 24 years. Of these, 52 isolates were collected between 2001 and 2009, which, before removing duplicates associated with outbreaks, represent 23% of the notified isolates over this period. To capture the diversity of human *E. coli* O157 strains observed in Australia and to avoid duplication, only single isolates were included from recognized outbreaks.

U.S. human isolates were selected from the Washington State University (WSU) culture collection to represent the most frequent XbaI PFGE

types isolated from human disease (with multiple isolates from outbreaks excluded) in the Centers for Disease Control and Prevention (CDC) PulseNet database. Cumulatively, these PFGE patterns represent those of 36.65% of all human cases in PulseNet since its inception. The isolates with these PFGE profiles were obtained from the Washington State Department of Health from among isolates obtained between 2004 and 2011. In addition, we included isolates representing 32 of the 39 SNP types described by Manning et al. (19), excluding only representatives of SNP types 11, 15, 17, 21, 22, 23, and 27 due to unavailability of viable pure cultures of those types.

LSPA-6 characterization. LSPA-6 analysis was performed using fluorescence-based capillary electrophoresis (27, 35). Targets were amplified with 6-carboxyfluorescein (FAM)- and VIC-labeled primers using the cycling conditions described by Yang et al. (35). Following amplification, products were serially diluted in sterile distilled water and separated by capillary electrophoresis using an Applied Biosystems 3130 Genetic Analyzer (Applied Biosystems) with a DS-33 matrix and GeneScan 600 LIZ size standard. Peak Scanner software (Version 1.0; Applied Biosystems) was used to interpret amplicon sizes, and LSPA-6 alleles were defined using the reference table provided by Yang et al. (35). In accordance with previous definitions, isolates shown to possess LSPA-6 genotype 111111 or 211111 were classified as lineage I (LI) or lineage I/II (LI/II), respectively, while all other allele combinations, including those with atypical sizes, were grouped as lineage II (LII) (35–37). Atypical alleles (those that could not be matched with previously published sizes) were assigned a new number in ascending order for each of the loci tested.

Analysis of SBI sites and Shiga toxin genes. A fluorescence-based docadecaplex PCR was used to determine the status of *stx* genes (*stx*₁, *stx*₂, and *stx*_{2c}) and bilateral phage-chromosome junctions for common SBI sites (*argW*, *sbcB*, *wrbA*, and *yehV*) in all isolates included in this study (38). This assay was designed to detect the presence of bacteriophage using primers specific for left and right bacteriophage junctions. In the absence of an inserted bacteriophage, the insertion site primers amplify the insertion site locus itself, distinguishable by size from a junction amplicon. Amplicons were separated using capillary electrophoresis and interpreted using Peak Scanner software (Version 1.0; Applied Biosystems, Foster City, CA). For ease of interpretation, SBI results were concatenated as follows. When either or both bacteriophage-chromosome insertion site locus junctions were detected, the locus was considered occupied. When an intact locus insertion site product was detected without amplification of either bacteriophage insertion site junction, the locus was considered unoccupied. A genotyping code was assigned to each isolate using the characters A, S, W, Y, 1, 2, and 2c to represent *argW*, *sbcB*, *wrbA*, *yehV*, *stx*₁, *stx*₂, and *stx*_{2c}, respectively. An N was assigned to any isolate that did not contain any of the three *stx* genes.

MLVA genotyping. MLVA was performed as outlined by Hyytiä-Trees et al. (39). Targets were analyzed in Australia using Peak Scanner software (Version 1.0; Applied Biosystems) and in the United States using GeneMarker software (Version 1.70; SoftGenetics, State College, PA). The number of allele repeats for each target was determined using a platform-specific reference table described by PulseNet (39). Partial fragments and fragment sizes differing from those listed in the reference table were confirmed by Sanger sequencing using an Applied Biosystems 3130 Genetic Analyzer (Applied Biosystems). MLVA repeats were uploaded into Bionumerics software (Version 4.61; Applied Maths, Austin, TX) as character data, and a minimum spanning tree was generated to assess the relationships between isolates of different origins. MLVA clusters were defined by isolates separated by no more than 2 allele differences and on the basis that each cluster contained at least 5 isolates and at least 5 nodes.

***tir* 255T>A polymorphism analysis.** A probe-based real-time PCR method was used to detect a single nucleotide polymorphism (A/T) located in the *tir* gene (21, 27).

Integrated analysis of SBI, LSPA-6, MLVA repeats, and *tir* 255T>A. Parsimony analysis using PAUP* (40) was chosen for the multivariate analysis because it was most capable of accurately dealing with the MLVA

TABLE 1 Distribution of LSPA-6 lineages, *tir* 255T>A polymorphisms, and *stx* genotypes in *E. coli* O157^a

Assay	Lineage, polymorphism, or genotype	Australia			United States			No. (%) of Australian isolates (n = 284)	No. (%) of U.S. isolates (n = 322)	P value
		No. (%) of cattle isolates (n = 205)	No. (%) of human isolates (n = 79)	P value	No. (%) of cattle isolates (n = 143)	No. (%) of human isolates (n = 179)	P value			
LSPA-6	LI	2 (1.0)	4 (5.1)	0.0523	65 (45.5)	127 (70.9)	<0.0001	6 (2.1)	192 (59.6)	<0.0001
	LI/II	165 (80.5)	75 (94.9)	0.0017	24 (16.8)	26 (14.5)	0.6430	240 (84.5)	50 (15.5)	<0.0001
	LII	38 (18.5)	0 (0.0)	<0.0001	54 (37.8)	26 (14.5)	<0.0001	38 (13.4)	80 (24.8)	0.0004
<i>tir</i> 255T>A	<i>tir</i> 255T	168 (82.0)	79 (100)	<0.0001	93 (65.0)	171 (95.5)	<0.0001	247 (87.0)	264 (82.0)	0.0946
	<i>tir</i> 255A	37 (18.0)	0 (0.0)	<0.0001	50 (35.0)	8 (4.5)	<0.0001	37 (13.0)	58 (18.0)	0.0946
<i>stx</i> analysis	<i>stx</i> ₁	4 (2.0)	6 (7.6)	0.0306	3 (2.1)	5 (2.8)	1.0000	10 (3.5)	8 (2.5)	0.3383
	<i>stx</i> ₂	1 (0.5)	1 (1.3)	0.4797	11 (7.7)	31 (17.3)	0.0122	2 (0.7)	42 (13.0)	<0.0001
	<i>stx</i> ₁ <i>stx</i> ₂	2 (1.0)	5 (6.3)	0.0194	63 (44.1)	114 (63.7)	0.0005	7 (2.5)	177 (55.0)	<0.0001
	<i>stx</i> ₁ <i>stx</i> ₂ <i>stx</i> _{2c}	0 (0.0)	0 (0.0)	1.0000	2 (1.4)	2 (1.1)	1.0000	0 (0.0)	4 (1.2)	0.1267
	<i>stx</i> ₁ <i>stx</i> _{2c}	123 (60.0)	52 (65.8)	0.3448	29 (20.3)	9 (5.0)	<0.0001	175 (61.6)	38 (11.8)	<0.0001
	<i>stx</i> _{2c}	74 (36.1)	14 (17.7)	0.0026	34 (23.8)	9 (5.0)	<0.0001	88 (31.0)	43 (13.4)	<0.0001
	<i>stx</i> ₂ <i>stx</i> _{2c}	0 (0.0)	1 (1.3)	0.4797	1 (0.7)	8 (4.5)	0.0470	1 (0.4)	9 (2.8)	0.0229
	N	1 (0.5)	0 (0.0)	1.0000	0 (0.0)	1 (0.6)	1.0000	1 (0.4)	1 (0.3)	1.0000

^a Number (percent) data represent isolates distributed by LSPA-6 (lineage-specific polymorphism assay), translocated intimin receptor (*tir*) 255T>A, and Shiga toxin (*stx*) genotypes. Bold data in P value columns represent significant differences ($P < 0.05$) between sources or countries within the corresponding row. Lineage I (LI) data refer to 111111, lineage I/II (LI/II) data refer to 211111, and lineage II (LII) data refer to all other combinations.

data. The MLVA repeats were encoded as single-digit alphanumeric characters (0 to 9 and A to Z), but PAUP* permitted the standard character set to be analyzed as ordinal values, which should provide a more accurate analysis than treating each encoded MLVA repeat as an independent value. This is based on the assumption that similar numbers of MLVA repeats reflect a higher degree of similarity between the isolates. A parsimony analysis was chosen to analyze the data as it was in character form. SBI data were encoded as 7 binary characters, LSPA-6 data were encoded as 6 characters with values in the range of 1 to 3, and *tir* 255T>A data were encoded as single binary characters. Character data for all taxa were converted to the Nexus format (41). The 609 taxa in the data set, including *E. coli* O157:H7 Sakai and *E. coli* O55 isolates (660-79 and 5905), were screened to remove all but one representative of those taxa with identical character sets. The total of 528 taxa retained for analysis represents a large number of taxa relative to a low number ($n = 22$) of parsimony-informative characters (0.042 characters per taxa), which might negatively affect the accuracy of phylogenetic analysis (42, 43). To maximize the accuracy of the analysis, a parsimony ratchet approach optimized for data sets with large numbers of taxa was used (44). This method has been demonstrated to be superior to the tree search methods typically employed by parsimony analysis programs (44, 45). Parsimony ratchet software generates a Nexus format script which directs the tree search methods of the data set in PAUP* (40). Testing determined that the optimal run length was 600 iterations with perturbation of one character. As mentioned above, the MLVA character set was analyzed as ordinal data whereas all other data were analyzed as nonordinal. Parsimony ratchet analysis was performed in 8 independent runs. The 12 shortest parsimony distance trees were selected from a representative run, and a majority rule tree was created with Dendroscope (46). A cluster network consensus generated in Dendroscope was visually identical to the majority rule tree with limited reticulation on minor branches indicating no significant changes in topology. Trees were formatted using FigTree v1.3.1 (47).

To further examine the clustering suggested by the parsimony analysis, the population structure of *E. coli* O157 was also inferred using STRUCTURE (version 2.2) software to model the multilocus genotype data from U.S. and Australian cattle and human isolates. This software uses a Bayesian clustering approach to detect the presence of population structure and to assign individuals probabilistically to populations while assuming that the genotyping loci are unlinked. While the software is

capable of using geographic or host source as prior information, it was applied in this instance without that information. Data were exported from the Nexus format into a comma-separated text file, retaining the coding system described in the previous section. All variables, including the MLVA data, were analyzed as nonordinal, due to the limitations of the software (27). Using software runs with 100,000 burn-in and 200,000 analysis iterations, we modeled cluster number (K) values ranging from 1 to 15. $K = 11$ was selected on the basis of a local maximum in posterior probability [$\ln P(X|K)$, where X represents the data]. All but 1 of 528 isolates with unique multilocus genotypes were assigned to identical clusters in duplicate independent model runs at $K = 11$. The only discrepantly clustered isolate had nearly equal (within 0.01) cluster assignment probabilities for the two clusters into which it was assigned in the two runs.

Statistical analyses. Statistical analyses using a 2-by-2 contingency table and Fisher's exact test (Minitab15; Minitab Inc., Minneapolis, MN) were performed on all values where the total value for the four quadrants was ≥ 40 and individual expected frequencies were as low as zero. P values were two-tailed, and groups were considered significantly different with P values < 0.05 .

RESULTS

Characterization of LSPA-6 genotypes. The distributions of LSPA-6 lineages in Australia and the United States were examined to gain insight into geographic structuring of *E. coli* O157 populations. A total of 26 unique lineage designations were obtained from the 606 isolates examined (see Table S1 in the supplemental material). Of these unique lineage designations, 23 were represented by U.S. isolates, while only 8 were found in isolates from Australia. In Australia, lineage LI/II represented 85% of all *E. coli* O157 isolates whereas LI and LII accounted for 2% and 13% of isolates, respectively (Table 1). In contrast, LI (60%) dominated among U.S. isolates, with smaller proportions of LII (25%) and LI/II (16%).

The host distributions of LI, LI/II, and LII followed similar patterns, albeit at greatly different frequencies, in the two countries. LI/II was similarly represented among isolates from cattle and humans (81% versus 95% in Australia and 17% versus 15% in

FIG 1 Distribution of Shiga toxin bacteriophage (SBI) genotypes in cattle and human *E. coli* O157 isolates sourced from Australia (AU; $n = 284$) and the United States (USA; $n = 322$). SBI characters A, S, W, and Y and 1, 2, and 2c were used to represent bacteriophage insertion sites *argW*, *sbcB*, *wrbA*, and *yehV* and *stx* genes *stx*₁, *stx*₂, and *stx*_{2c}, respectively. SBI profiles that occurred in less than 10 isolates were grouped as “other,” and complete SBI data are presented in Table S2 in the supplemental material.

the United States), LII isolates were significantly ($P < 0.05$) more frequent among cattle than humans (19% versus 0% in Australia and 38% versus 15% in the United States), and LI isolates were more frequent among humans than cattle (5% versus 1% in Australia and 71% versus 46% in the United States).

Identification of SBI profiles. SBI typing characterizes *E. coli* O157 strain genotypes based on the detection of Stx-encoding bacteriophage-chromosome insertion site junctions and specific Stx genes by PCR, and SBI genotypes are reported as the concatenated abbreviated Stx-encoding bacteriophage insertion locus followed by the concatenated Stx subtype designation(s). To establish whether *E. coli* O157 isolates from Australia and the United States could be differentiated on the basis of SBI, we interrogated 606 isolates for *stx* genes and chromosomal loci commonly known to harbor *stx* bacteriophages (Fig. 1). Of the 606 isolates tested, 31 unique SBI genotypes were identified, and 13 of the 31 represented single isolates (see Table S2 in the supplemental material). Overall, isolates from the United States demonstrated greater SBI diversity than Australian isolates, and different genotypes predominated in each region. Genotype ASY12c predominated among Australian isolates (49%) but was significantly less frequent among U.S. isolates (7%) ($P < 0.05$). In contrast, genotype WY12 accounted for significantly ($P < 0.05$) more U.S. isolates (55%) than Australian isolates (2%). SBI types SY2c (31%) and AS12c (11%) represented the second and third most frequent genotypes in Australia and, as with ASY12c, occurred in significantly more Australian isolates than U.S. isolates ($P < 0.05$).

Not surprisingly, carriage of Shiga toxin genes also varied among isolates by country of origin (Table 1). The occurrence of *stx*₂ in U.S. isolates (72%) was significantly greater than the occurrence of *stx*₂ in Australian isolates (4%) ($P < 0.05$), while the opposite was true for *stx*_{2c} which was represented in a greater

proportion of Australian isolates (93%) than U.S. isolates (29%) ($P < 0.05$). In contrast, the prevalences of *stx*₁ did not differ significantly between Australia (68%) and the United States (70%). U.S. isolates were dominated by four main *stx* genotypes, *stx*₁ *stx*₂ (55%), *stx*_{2c} (13%), *stx*₂ (13%), and *stx*₁ *stx*_{2c} (12%), while Australian isolates were dominated by two, *stx*₁ *stx*_{2c} (62%) and *stx*_{2c} (31%). All other *stx* genotypes were represented in less than 5% of isolates in each country.

The distribution of SBI genotypes in human and cattle isolates was assessed to determine if the trend observed between countries was also present between isolate sources (see Table S2 in the supplemental material). A greater proportion of U.S. human isolates than cattle isolates were represented by WY12 ($P < 0.05$), and while this trend was also observed in Australia, the difference was not statistically significant ($P = 0.0526$). In both Australia and the United States, SY2c occurred in a greater proportion of cattle isolates than human isolates ($P < 0.05$). Similarly, ASY12c occurred in a greater proportion of U.S. cattle isolates than human isolates ($P < 0.05$); in Australian isolates, however, this trend was not statistically significant ($P > 0.05$).

MLVA. Characterization of isolates by MLVA determined that 518 of 606 isolates possessed unique MLVA types. No identical MLVA profiles were common between the two countries (Fig. 2A). In a single instance, a U.S. human isolate shared a single profile with a U.S. cattle isolate. In all other cases, indistinguishable profiles were observed only between isolates from the same country and the same source. A total of 13 MLVA clusters were identified on the basis that a cluster contained at least 5 isolates and 5 nodes and that isolates within a cluster were separated by no more than 2 allele differences (Fig. 2B). All isolates within clusters 1 to 6, 8, and 10 were derived from the United States, while clusters 7, 9, 12, and 13 contained only Australian isolates. A single cluster

FIG 2 Minimum spanning tree of 609 isolates (including *E. coli* O157 Sakai and 2 *E. coli* O55 isolates) examined using multilocus variable-number tandem-repeat analysis (MLVA). Each node represents a unique MLVA type. (A) Comparison based on country and source. Divisions in nodes represent the number of isolates with that MLVA profile. Isolates are color coded by country and source as follows: Australian cattle, light orange; Australian humans, dark orange; U.S. cattle, light purple; U.S. humans, dark purple; *E. coli* O157:H7 Sakai, red; and *E. coli* O55, green. (B) Clustering of MLVA profiles. The number of isolates within each node is reflected by the size of the node. Clusters are color coded and are defined by isolates separated by no more than 2 allele differences and on the basis that each cluster contained at least 5 isolates and at least 5 nodes. Each cluster has been labeled with a number from 1 to 13.

(cluster 11) contained isolates from both the United States (cattle, $n = 10$; human, $n = 2$) and Australia (cattle, $n = 1$; human, $n = 0$).

***tir* 255T>A polymorphism analysis.** We investigated the presence of a polymorphism in the *tir* gene, previously used to identify human- and cattle-biased groups, to determine its significance to Australian and U.S. cattle and human isolates. Overall, 87% of Australian isolates and 82% of U.S. isolates were shown to possess the *tir* 255T allele (Table 1). In both countries, *tir* 255A was more likely to occur in cattle isolates than human isolates (18% versus 0% in Australia and 35% versus 5% in the United States), while the opposite was true for *tir* 255T, which was more frequent among isolates from humans than in those from cattle (100%

versus 82% in Australia and 96% versus 65% in the United States) ($P < 0.05$).

Comparison of genotyping methods. Further analysis was performed to determine the relationships among the various genotyping methods. SBI types comprising 5 or more isolates collectively represented 95% of all isolates. Of these, genotypes WY12 (95%), W2 (89%), and WY2 (80%) were predominantly represented by LI ($P < 0.05$) (see Table S3 in the supplemental material). Similarly, AS12c (100%), ASY12c (93%), AY2 (65%), ASY22c (80%), and A1 (80%) were significantly associated with LI/II ($P < 0.05$). Isolates that possessed genotype SY2c were mostly distributed between LII (54%) and LI/II (46%), with LI accounting for only 1% of isolates.

A strong association between MLVA clusters and LSPA-6 genotypes was also observed. Clusters 1 to 4 contained a combined 108 isolates, 97% of which were LI. Similarly, clusters 5, 6, 8, and 10 together represented 60 isolates, 96% of which were LII, and clusters 9, 11, and 12 to 13 comprised 68 isolates, 99% of which were LI/II. In contrast, a single cluster (cluster 7) representing 63 isolates was associated with both LII (42%) and LI/II (62%).

Further associations were observed between *tir* and LSPA-6 lineage (see Table S4 in the supplemental material) and between *tir*, LSPA-6, and SBI (Table S3). The *tir* 255T SNP was carried by more than 99% of LI isolates ($n = 198$) and LI/II isolates ($n = 290$). In contrast to this, LII isolates ($n = 118$) were more likely to be represented by *tir* 255A (78%) than *tir* 255T (22%). In addition, eight of the 10 most common SBI genotypes displayed a greater than 90% association with *tir* 255T (Table S3).

Integrated analysis of SBI, LSPA-6, MLVA, and *tir* 255T>A data. Maximum parsimony analysis was used to construct a consensus tree based on the data from all of the genotyping methods to further investigate the relationships among isolates from different countries and sources (Fig. 3). The resultant tree demonstrates a high degree of segregation among the isolates based on the country of origin. Each of the six parsimony groups was significantly associated with a country of origin: the majority of Australian isolates (84%) segregated into two parsimony groups (PG4 and PG6), while most of the U.S. isolates (70%) segregated into parsimony groups PG1, PG2, and PG5. Generally, there was a lack of separate clustering for Australian cattle and human isolates compared to the distinct clustering of U.S. cattle and human isolates. Specifically, the parsimony group (PG6) that contained the majority of Australian isolates (72%) was not overrepresented by either human or cattle isolates; however, the second-most-populated group (PG4) was significantly associated with cattle. In contrast, two parsimony groups (PG1 and PG2) were significantly associated with U.S. human isolates and one (PG5) was significantly associated with U.S. cattle isolates (Table 2). In an alternative approach to analysis of the multilocus genotyping data, a Bayesian technique was used to probabilistically assign isolates to 11 genotype clusters that were similarly strongly associated by country of origin, host source, or both. Of 9 clusters to which 25 or more isolates were assigned, 9 were strongly associated with a U.S. origin (3 clusters) or an Australian origin (6 clusters) (Table 3). Cluster McMC-V alone included a significant percentage of isolates from both countries (13% Australia, 19% United States) and was strongly associated with cattle in both source countries.

FIG 3 Majority consensus tree of the 12 lowest-scoring trees from the ratchet parsimony analysis of 528 unique isolates. Taxa are coded by shape and color according to the country and source as follows: Australian cattle, light orange circles; Australian humans, dark orange circles; U.S. cattle, light purple squares; U.S. humans, dark purple squares; *E. coli* O157:H7 Sakai, red diamond; and *E. coli* O55 isolates 660-79 and 5905, green triangles. Taxon markers were manually rearranged on the branches to reveal those markers that were hidden by overlapping present in the initial computer output. Markers representing two or more isolates with identical multilocus genotypes are scaled accordingly, with the number of taxa (where >1) represented in the center of the marker. Parsimony groups are indicated by dashed lines around the major branches in the tree.

DISCUSSION

Genotypic differences among *E. coli* O157 isolates from Australia and other countries were first reported in 2001 by Kim et al. (32) using octamer-based genome scanning (OBGS). These findings were later reinforced by studies examining chromosomal gene polymorphisms and Shiga toxin genotypes (34, 48, 49). However, the limited strain sets used in these studies make it difficult to draw significant conclusions about geographic diversity. There is also

some evidence to support the idea of divergent populations of *E. coli* O157, with the predominance of particular subtypes observed in other countries such as The Netherlands (50, 51) and Japan (50, 51). The goal of this study was to identify genotypic differences between larger sets of representative *E. coli* O157 isolates from diverse geographic regions such as Australia and the United States. We hypothesized that *E. coli* O157 from these two countries would be differentiated based on LSPA-6, *tir* 255T>A, SBI, and MLVA typing, and the results of this analysis have supported this hypothesis.

The lineage-specific polymorphism assay was first developed by Yang et al. (35) to separate *E. coli* O157 isolates into two genotypes referred to as “lineages” (LI and LII) on the basis of observed differences in six loci representing conserved regions of the O157 backbone. The assay has since been reinterpreted to define a third lineage (LI/II). The various frequencies at which these lineages are isolated from cattle and humans suggests that the LSPA-6 assay may be a useful indicator of virulence potential. Assessment of 606 isolates by LSPA-6 analysis revealed a striking difference between the Australian and U.S. isolates. In the United States, LI was the most frequently detected genotype in both cattle and clinical isolates, while Australian isolates were dominated by LI/II. LI isolates have been associated with clinical infection in many countries, including the United States, Japan, and Canada, and LI isolates in this study occurred in significantly more human isolates than cattle isolates. LI/II has also been associated with clinical disease, but it has recently become apparent that certain clades within LI/II are responsible for more frequent and severe disease than others (19). Our recent geographic comparison of Argentinean and Australian *E. coli* O157 isolates identified LI/II as the predominant lineage in both countries (34); however, further analysis of virulence clades according to Manning et al. (19) distinguished LI/II Australian isolates (clade 7) from LI/II Argentinean isolates (clade 8) (34). Interestingly, isolates belonging to clade 8 have been associated with more-severe disease outcomes, including HUS, while those belonging to clade 7 are associated with less-severe disease (19). Clade 7 has also been described as a bovine-biased group, further supporting previous claims that this genotype is of limited clinical significance relative to clade 8 (52). Bono et al. (26) used nucleotide polymorphisms to classify human- and cattle-derived *E. coli* O157 isolates into eight major genotype lineages, seven of which were found in cattle. Two of these seven cattle lineages also contained the majority of human isolates, reinforcing the identifica-

TABLE 2 Parsimony groupings of *E. coli* O157 based on MLVA, LSPA-6, *tir* 255T>A, and SBI^a

Parsimony group	Australia			United States			No. (%) of Australian isolates (n = 284)	No. (%) of U.S. isolates (n = 322)	P value
	No. (%) of cattle isolates (n = 205)	No. (%) of human isolates (n = 79)	P value	No. (%) of cattle isolates (n = 143)	No. (%) of human isolates (n = 179)	P value			
PG1	0 (0.0)	3 (3.8)	0.0209	42 (29.4)	75 (41.9)	0.0265	3 (1.1)	117 (36.3)	<0.0001
PG2	0 (0.0)	0 (0.0)	1.0000	14 (9.8)	33 (18.4)	0.0382	0 (0.0)	47 (14.6)	<0.0001
PG3	1 (0.5)	1 (1.3)	0.4797	8 (5.6)	17 (9.5)	0.2149	2 (0.7)	25 (7.8)	<0.0001
PG4	31 (15.1)	3 (3.8)	0.0074	12 (8.4)	7 (3.9)	0.1005	34 (12.0)	19 (5.9)	0.0094
PG5	11 (5.4)	1 (1.3)	0.1891	47 (32.9)	14 (7.8)	<0.0001	12 (4.2)	61 (18.9)	<0.0001
PG6	144 (70.2)	60 (75.9)	0.3792	12 (8.4)	6 (3.4)	0.0849	204 (71.8)	18 (5.6)	<0.0001
N ^b	18 (8.8)	11 (13.9)	0.1977	8 (5.6)	27 (15.1)	0.0067	29 (10.2)	35 (10.9)	0.8948

^a Number (percent) data represent isolates distributed by parsimony group. Bold data in P value columns represent significant differences ($P < 0.05$) between sources or countries within the corresponding row.

^b N, isolates that were not located in any of the six parsimony groups defined in the Fig. 3 legend.

TABLE 3 Markov chain Monte Carlo analysis of *E. coli* O157 based on LSPA-6, *tir* 255T>A, SBI, and MLVA^a

Cluster	Total no. of isolates	Australia			United States			No. (%) of Australian isolates (n = 284)	No. (%) of U.S. isolates (n = 322)	P value
		No. (%) of cattle isolates (n = 205)	No. (%) of human isolates (n = 79)	P value	No. (%) of cattle isolates (n = 143)	No. (%) of human isolates (n = 179)	P value			
McMC-I	9	0 (0.0)	0 (0.0)	1.0000	6 (4.2)	3 (1.7)	0.1929	0 (0.0)	9 (2.8)	0.0042
McMC-II	198	2 (1.0)	4 (5.1)	0.0526	62 (43.4)	130 (72.6)	<0.0001	6 (2.1)	192 (59.6)	<0.0001
McMC-III	31	2 (1.0)	2 (2.5)	0.3095	8 (5.6)	19 (10.6)	0.1556	4 (1.4)	27 (8.4)	<0.0001
McMC-IV	38	22 (10.7)	6 (7.6)	0.5105	4 (2.8)	6 (3.4)	1.0000	28 (9.9)	10 (3.1)	0.0007
McMC-V	97	36 (17.6)	0 (0.0)	<0.0001	50 (35.0)	11 (6.1)	0.0009	36 (12.7)	61 (18.9)	0.0453
McMC-VI	25	8 (3.9)	17 (21.5)	<0.0001	0 (0.0)	0 (0.0)	1.0000	25 (8.8)	0 (0.0)	<0.0001
McMC-VII	42	16 (7.8)	25 (31.6)	<0.0001	0 (0.0)	1 (0.6)	1.0000	41 (14.4)	1 (0.3)	<0.0001
McMC-VIII	35	31 (15.1)	3 (3.8)	0.0047	0 (0.0)	1 (0.6)	1.0000	34 (12.0)	1 (0.3)	<0.0001
McMC-IX	25	8 (3.9)	10 (12.7)	0.0121	2 (1.4)	5 (2.8)	0.4687	18 (6.3)	7 (2.2)	0.0130
McMC-X	23	6 (2.9)	3 (3.8)	0.7122	11 (7.7)	3 (1.7)	0.0116	9 (3.2)	14 (4.3)	0.526
McMC-XI	83	74 (36.1)	9 (11.4)	<0.0001	0 (0.0)	0 (0.0)	1.0000	83 (29.2)	0 (0.0)	<0.0001

^a Number (percent) data represent isolates distributed by Markov chain Monte Carlo (MCMC) group. Bold data in P value columns represent significant differences ($P < 0.05$) between sources or countries within the corresponding row.

tion of cattle as a reservoir of clinically significant *E. coli* O157 strains. However, one lineage containing cattle isolates was under-represented by those from humans, and the authors concluded that this lineage may have evolved away from causing human disease (26). This further supports the growing evidence for bovine-biased groups of *E. coli* O157 strains which may have reduced disease potential in humans.

Specific clades and lineages of *E. coli* O157 have been associated with particular *stx*-prophage combinations (19, 34). Results from this study further indicate that *stx* prophages are associated with particular lineages. For example, our data demonstrate that LI isolates, most frequently SBI genotype WY12, are rarely associated with phage insertions in *argW* or *sbcB* or with *stx*_{2c}. LI/II isolates commonly have phage inserted in *argW*, *sbcB*, and *yehV*, encoding Stx subtypes Stx1, Stx2, and Stx2c, but infrequently contain phage inserted in *wrbA*. In this study, we observed that Australian isolates carrying *stx*₁ also frequently carry an occupied *argW* locus, suggesting that the *stx*₁-prophage combination in these strains is inserted at this locus. This observation is consistent with our previous study that demonstrated that the integration of an *stx*₁-prophage combination at *argW* is characteristic of Australian isolates (34). While a small percentage of U.S. isolates also had an SBI profile (ASY12c) consistent with an *stx*₁-prophage insertion in *argW*, the association of *stx*₁ with this locus has yet to be unequivocally demonstrated in these strains. In light of the apparent stable association of particular *stx*-prophage combinations with LSPA-6 lineages, we suggest that *stx*-prophage combinations may block the integration of other phage in the occupied locus and may also contribute to the exclusion of additional *stx*-prophage combinations in additional unoccupied loci.

The carriage of *stx* subtypes (*stx*₁, *stx*₂, and *stx*_{2c}) has previously been suggested to influence the pathogenic potential of *E. coli* O157 (53). A prior study comparing Stx toxicities revealed that Stx2 was 40 to 400 times more potent against mice than Stx1 or Stx2c (53). Likewise, the effect of different *stx*-bacteriophage combinations (e.g., *stx*₂-*wrbA* versus *stx*₂-*argW* versus *stx*₁-*argW*) on the expression of Stx may also contribute to the virulence potential of an isolate. Here, we demonstrated that the prevalences of *stx*₂ and *stx*_{2c} differ substantially in isolates from Australia and the United States. Isolates harboring *stx*₂ were frequently present in

the United States (72%) but rarely present in Australia (4%), while those carrying *stx*_{2c} were more frequent in Australia (93%) than in the United States (29%). This study also demonstrated that the frequency of *stx*₂ was high in LI isolates but low among LI/II and LII isolates.

Genotypic diversity between isolates derived from bovine and human origins based on gene or single nucleotide polymorphisms has previously been identified (18, 21, 54), and this work supports these findings. Studies have shown an association between a specific *tir* polymorphism and the likelihood of isolation from cattle or humans (21, 26). Isolates harboring the *tir* 255A SNP are rarely isolated from human clinical cases but frequently isolated from cattle. This finding is supported by the results of the current study. Nevertheless, isolates possessing the *tir* 255T comprise substantial percentages of cattle isolates from both Australia and the United States. Comparisons of lineage and *tir* SNP data suggest that *tir* 255A is generally absent from LI or LI/II isolates in both Australia and the United States. Additionally, this study did not reveal a significant association between LI/II isolates and hosts but did identify associations between LI and LII with human and cattle isolates, respectively. This suggests that *tir* 255A may be highly correlated with LII isolates and by itself is a useful predictor of lineage as opposed to a predictor of source (cattle or human) or pathogenic potential.

MLVA typing results further support the idea of the geographic separation of *E. coli* O157 strains. Isolates from both countries largely grouped into distinct branches, and no Australian and U.S. isolates shared identical MLVA patterns. This result may reflect the inherent nature of MLVA in providing highly discriminatory information about *E. coli* O157 genotypes that is useful in epidemiological investigations.

Consistent with the results obtained with each individual genotyping system, multilocus analyses of the entire genotyping data set (SBI, LSPA-6, MLVA, and *tir* 255T>A) using maximum parsimony and of a subsection of the data (SBI, LSPA-6, and *tir* 255T>A) by the Markov chain Monte Carlo (MCMC) method offer strong evidence of geographic structuring of *E. coli* O157 populations, suggesting divergent evolution of enterohemorrhagic *E. coli* O157 in Australia and the United States. Maximum parsimony analysis showed that the majority of Australian isolates

(PG4 and PG6) were closer to branches overrepresented by U.S. cattle isolates (PG5) than to branches containing a higher proportion of U.S. human isolates (PG1, PG2, and PG3).

The incidences and severities of disease caused by *E. coli* O157 differ substantially worldwide (9–13). In comparison to many countries, including the United States, Australia has a low reported incidence of infection (9). Food-handling practices, consumer preferences for foods, environmental factors, and host susceptibility are among the many factors that may influence the likelihood of infection and account for the regional discrepancies in disease characteristics. However, microbial genetic factors are gaining increased recognition as important determinants of human infection risk (18, 19, 21, 36). This is supported by growing epidemiological evidence of the existence of clinically significant genotypes. There is also evidence of substantial variability in the virulence of *E. coli* O157 genotypes implicated in different outbreaks (19, 22, 23) and that patient symptoms vary with respect to the genotype of *E. coli* O157 involved (19). It is therefore reasonable to propose that countries with high rates of disease may be populated with more clinically significant genotypes of *E. coli* O157 than countries with low rates of disease. The geographic divergence and structuring identified in this study may be among the factors associated with the differences in the incidences of *E. coli* O157-associated human disease in Australia and the United States. Although the specific genetic factors underlying such differential virulence levels cannot be determined from our results, some candidate genes which may play a role are represented by the *stx*₁-*argW* carriage uniquely observed in Australian isolates (associated with low virulence) and the much higher frequency of *stx*₂ carriage (associated with higher virulence) in strains in the United States. Nevertheless, it is possible that additional unmeasured factors contribute to differential virulence levels and that all markers used in this study simply identify the lineages that carry unknown virulence factors. The potential for geographic structuring of other bacterial species within animal populations, e.g., *Salmonella* subsp. II, serotype Sofia, which is commonly isolated from Australian broiler chickens (55) but has rarely been associated with poultry in other parts of the globe, has also been observed.

In conclusion, it is clear from this work that the U.S. and Australian *E. coli* O157 populations have diverged based on geographical isolation. It is possible that the Australian population of *E. coli* O157 strains may have emerged from a bovine-biased subpopulation which has become widely distributed in Australia and is therefore the major genetic type associated with clinical disease in this country. In contrast, the *E. coli* O157 population in the United States appears to be more genetically diverse across both human and bovine sources. This work also suggests that this geographical bias in *E. coli* O157 populations may be associated with the human epidemiology of this pathogen and that more detailed investigations into geographical population structuring and associated virulence are therefore warranted.

ACKNOWLEDGMENTS

We acknowledge Simon Gladman (CSIRO Animal, Food and Health Sciences) for his assistance in providing custom scripts used to generate the Nexus data file and Chad Laing (Public Health Agency of Canada) for his assistance in formatting data for parsimony ratchet and STRUCTURE analyses. We also acknowledge the expert technical assistance of Katie Baker and Smriti Shringi.

We acknowledge support provided by USDA AFRI grants 2009-04248

and 2010-04487. We also gratefully acknowledge funding from Meat & Livestock Australia and supporting funds provided by the Commonwealth Scientific and Industrial Research Organization (CSIRO) and the Department of Primary Industries Victoria.

REFERENCES

- García A, Fox JG, Besser TE. 2010. Zoonotic enterohemorrhagic *Escherichia coli*: a One Health perspective. *ILAR J.* 51:221–232.
- Elder RO, Keen JE, Siragusa GR, Barkocy-Gallagher GA, Koohmaraie M, Laegreid WW. 2000. Correlation of enterohemorrhagic *Escherichia coli* O157 prevalence in feces, hides, and carcasses of beef cattle during processing. *Proc. Natl. Acad. Sci. U. S. A.* 97:2999–3003.
- Ferens WA, Hovde CJ. 2011. *Escherichia coli* O157:H7: animal reservoir and sources of human infection. *Foodborne Pathog. Dis.* 8:465–487.
- Scallan E, Hoekstra RM, Angulo FJ, Tauxe RV, Widdowson MA, Roy SL, Jones JL, Griffin PM. 2011. Foodborne illness acquired in the United States—major pathogens. *Emerg. Infect. Dis.* 17:7–15.
- Frenzen PD, Drake A, Angulo FJ. 2005. Economic cost of illness due to *Escherichia coli* O157 infections in the United States. *J. Food Prot.* 68:2623–2630.
- Masana MO, Leotta GA, Del Castillo LL, D'Astak BA, Palladino PM, Galli L, Vilacoba E, Carbonari C, Rodriguez HR, Rivas M. 2010. Prevalence, characterization, and genotypic analysis of *Escherichia coli* O157:H7/NM from selected beef exporting abattoirs of Argentina. *J. Food Prot.* 73:649–656.
- Pollock KGJ, Locking ME, Beattie TJ, Maxwell H, Ramage I, Hughes D, Cowieson J, Allison L, Hanson M, Cowden JM. 2010. Sorbitol-fermenting *Escherichia coli* O157, Scotland. *Emerg. Infect. Dis.* 16:881–882.
- Pollock KGJ, Young D, Beattie TJ, Todd WTA. 2008. Clinical surveillance of thrombotic microangiopathies in Scotland, 2003–2005. *Epidemiol. Infect.* 136:115–121.
- Vally H, Hall G, Dyda A, Raupach J, Knope K, Combs B, Desmarchelier P. 21 January 2012. Epidemiology of Shiga toxin producing *Escherichia coli* in Australia, 2000–2010. *BMC Public Health* 12:63. doi:10.1186/1471-2458-12-63.
- Locking ME, Pollock KGJ, Allison LJ, Rae L, Hanson MF, Cowden JM. 2011. *Escherichia coli* O157 infection and secondary spread, Scotland, 1999–2008. *Emerg. Infect. Dis.* 17:524–527.
- Centers for Disease Control Prevention. 2011. Vital signs: incidence and trends of infection with pathogens transmitted commonly through food—foodborne diseases active surveillance network, 10 U.S. sites, 1996–2010. *MMWR Morb. Mortal. Wkly. Rep.* 60:749–755.
- Rivas M, Miliwebsky ES, Chinen I, Deza N, Leotta GA. 2006. Epidemiology of the hemolytic uremic syndrome in Argentina. Diagnosis of the etiological agent, reservoirs and routes of transmission. *Medicina (B. Aires)* 66(Suppl 3):27–32. (In Spanish.)
- Tarr PI, Gordon CA, Chandler WL. 2005. Shiga-toxin-producing *Escherichia coli* and haemolytic uraemic syndrome. *Lancet* 365:1073–1086.
- Fernández D, Rodríguez EM, Arroyo GH, Padola NL, Parma AE. 2009. Seasonal variation of Shiga toxin-encoding genes (*stx*) and detection of *E. coli* O157 in dairy cattle from Argentina. *J. Appl. Microbiol.* 106:1260–1267.
- Omisakin F, MacRae M, Ogden ID, Strachan NJC. 2003. Concentration and prevalence of *Escherichia coli* O157 in cattle feces at slaughter. *Appl. Environ. Microbiol.* 69:2444–2447.
- Fegan N, Vanderlinde P, Higgs G, Desmarchelier P. 2004. The prevalence and concentration of *Escherichia coli* O157 in faeces of cattle from different production systems at slaughter. *J. Appl. Microbiol.* 97:362–370.
- Barlow RS, Mellor GE. 2010. Prevalence of enterohemorrhagic *Escherichia coli* serotypes in Australian beef cattle. *Foodborne Pathog. Dis.* 7:1239–1245.
- Shringi S, García A, Lahmers KK, Potter KA, Muthupalani S, Swennes AG, Hovde CJ, Call DR, Fox JG, Besser TE. 2012. Differential virulence of clinical and bovine-biased enterohemorrhagic *Escherichia coli* O157:H7 genotypes in piglet and Dutch belted rabbit models. *Infect. Immun.* 80:369–380.
- Manning SD, Motiwala AS, Springman AC, Qi W, Lacher DW, Ouellette LM, Mladonicky JM, Somsel P, Rudrik JT, Dietrich SE, Zhang W, Swaminathan B, Alland D, Whittam TS. 2008. Variation in virulence among clades of *Escherichia coli* O157:H7 associated with disease outbreaks. *Proc. Natl. Acad. Sci. U. S. A.* 105:4868–4873.

20. Zhang Y, Laing C, Zhang Z, Hallewell J, You C, Ziebell K, Johnson RP, Kropinski AM, Thomas JE, Karmali M, Gannon VP. 2010. Lineage and host source are both correlated with levels of Shiga toxin 2 production by *Escherichia coli* O157:H7 strains. *Appl. Environ. Microbiol.* 76:474–482.
21. Bono JL, Keen JE, Clawson ML, Durso LM, Heaton MP, Laegreid WW. 2007. Association of *Escherichia coli* O157:H7 *tir* polymorphisms with human infection. *BMC Infect. Dis.* 7:98. doi:10.1186/1471-2334-7-98.
22. Fukushima H, Hashizume T, Morita Y, Tanaka J, Azuma K, Mizumoto Y, Kaneno M, Matsuura M, Konma K, Kitani T. 1999. Clinical experiences in Sakai City Hospital during the massive outbreak of enterohemorrhagic *Escherichia coli* O157 infections in Sakai City, 1996. *Pediatr. Int.* 41:213–217.
23. Grant J, Wendelboe AM, Wendel A, Jepson B, Torres P, Smelser C, Rolfs RT. 2008. Spinach-associated *Escherichia coli* O157:H7 outbreak, Utah and New Mexico, 2006. *Emerg. Infect. Dis.* 14:1633–1636.
24. Ziebell K, Steele M, Zhang Y, Benson A, Taboada EN, Laing C, McEwen S, Ciebin B, Johnson R, Gannon V. 2008. Genotypic characterization and prevalence of virulence factors among Canadian *Escherichia coli* O157:H7 strains. *Appl. Environ. Microbiol.* 74:4314–4323.
25. Yokoyama E, Hashimoto R, Etoh Y, Ichihara S, Horikawa K, Uchimura M. 2011. Biased distribution of IS629 among strains in different lineages of enterohemorrhagic *Escherichia coli* serovar O157. *Infect. Genet. Evol.* 11:78–82.
26. Bono JL, Smith TP, Keen JE, Harhay GP, McDanel TG, Mandrell RE, Jung WK, Besser TE, Gerner-Smidt P, Bielaszewska M, Karch H, Clawson ML. 2012. Phylogeny of Shiga toxin-producing *Escherichia coli* O157 isolated from cattle and clinically ill humans. *Mol. Biol. Evol.* 29:2047–2062.
27. Whitworth J, Zhang Y, Bono J, Pleydell E, French N, Besser T. 2010. Diverse genetic markers concordantly identify bovine origin *Escherichia coli* O157 genotypes underrepresented in human disease. *Appl. Environ. Microbiol.* 76:361–365.
28. Feng P, Lampel KA, Karch H, Whittam TS. 1998. Genotypic and phenotypic changes in the emergence of *Escherichia coli* O157:H7. *J. Infect. Dis.* 177:1750–1753.
29. Feng PCH, Monday SR, Lacher DW, Allison L, Siitonen A, Keys C, Eklund M, Nagano H, Karch H, Keen J, Whittam TS. 2007. Genetic diversity among clonal lineages within *Escherichia coli* O157: H7 stepwise evolutionary model. *Emerg. Infect. Dis.* 13:1701–1706.
30. Zhou Z, Li X, Liu B, Beutin L, Xu J, Ren Y, Feng L, Lan R, Reeves PR, Wang L. 2010. Derivation of *Escherichia coli* O157:H7 from its O55:H7 precursor. *PLoS One* 5:e8700. doi:10.1371/journal.pone.0008700.
31. Leopold SR, Magrini V, Holt NJ, Shaikh N, Mardis ER, Cagno J, Ogura Y, Iguchi A, Hayashi T, Mellmann A, Karch H, Besser TE, Sawyer SA, Whittam TS, Tarr PI. 2009. A precise reconstruction of the emergence and constrained radiations of *Escherichia coli* O157 portrayed by backbone concatenomic analysis. *Proc. Natl. Acad. Sci. U. S. A.* 106:8713–8718.
32. Kim J, Nietfeldt J, Ju J, Wise J, Fegan N, Desmarchelier P, Benson AK. 2001. Ancestral divergence, genome diversification, and phylogeographic variation in subpopulations of sorbitol-negative, beta-glucuronidase-negative enterohemorrhagic *Escherichia coli* O157. *J. Bacteriol.* 183:6885–6897.
33. Wick LM, Qi W, Lacher DW, Whittam TS. 2005. Evolution of genomic content in the stepwise emergence of *Escherichia coli* O157:H7. *J. Bacteriol.* 187:1783–1791.
34. Mellor GE, Sim EM, Barlow RS, D'Astak BA, Galli L, Chinen I, Rivas M, Gobius KS. 2012. Phylogenetically related Argentinean and Australian *Escherichia coli* O157 isolates are distinguished by virulence clades and alternative Shiga toxin 1 and 2 prophages. *Appl. Environ. Microbiol.* 78:4724–4731.
35. Yang Z, Kovar J, Kim J, Nietfeldt J, Smith DR, Moxley RA, Olson ME, Fey PD, Benson AK. 2004. Identification of common subpopulations of non-sorbitol-fermenting, beta-glucuronidase-negative *Escherichia coli* O157:H7 from bovine production environments and human clinical samples. *Appl. Environ. Microbiol.* 70:6846–6854.
36. Zhang Y, Laing C, Steele M, Ziebell K, Johnson R, Benson AK, Taboada E, Gannon VPJ. 16 May 2007. Genome evolution in major *Escherichia coli* O157:H7 lineages. *BMC Genomics* 8:121. doi:10.1186/1471-2164-8-121.
37. Steele M, Ziebell K, Zhang YX, Benson A, Konczyk P, Johnson R, Gannon V. 2007. Identification of *Escherichia coli* O157: H7 genomic regions conserved in strains with a genotype associated with human infection. *Appl. Environ. Microbiol.* 73:22–31.
38. Shringi S, Schmidt C, Katherine K, Brayton KA, Hancock DD, Besser TE. 2012. Carriage of *stx2a* differentiates clinical and bovine-biased strains of *Escherichia coli* O157. *PLoS One* 7:e51572. doi:10.1371/journal.pone.0051572.
39. Hyttia-Trees E, Lafon P, Vauterin P, Ribot EM. 2010. Multilaboratory validation study of standardized multiple-locus variable-number tandem repeat analysis protocol for Shiga toxin-producing *Escherichia coli* O157: a novel approach to normalize fragment size data between capillary electrophoresis platforms. *Foodborne Pathog. Dis.* 7:129–136.
40. Wilgenbusch JC, Swofford D. 2003. Inferring evolutionary trees with PAUP*. *Curr. Protoc. Bioinformatics* 2003:6.4.1–6.4.28.
41. Maddison DR, Swofford DL, Maddison WP. 1997. NEXUS: an extensible file format for systematic information. *Syst. Biol.* 46:590–621.
42. Charleston MA, Hendy MD, Penny D. 1994. The effects of sequence length, tree topology, and number of taxa on the performance of phylogenetic methods. *J. Comput. Biol.* 1:133–151.
43. Hillis DM, Huelsenbeck JP, Cunningham CW. 1994. Application and accuracy of molecular phylogenies. *Science* 264:671–677.
44. Nixon KC. 1999. The parsimony ratchet, a new method for rapid parsimony analysis. *Cladistics* 15:407–414.
45. Goloboff PA. 1999. Analyzing large data sets in reasonable times: solutions for composite optima. *Cladistics* 15:415–428.
46. Huson DH, Richter DC, Rausch C, DeZulian T, Franz M, Rupp R. 2007. Dendroscope: an interactive viewer for large phylogenetic trees. *BMC Bioinformatics* 8:460. doi:10.1186/1471-2105-8-460.
47. Rambaut A. 2010, posting date. FigTree 1.3.1. <http://tree.bio.ed.ac.uk/software/figtree/>.
48. Yang ZJ, Kim J, Zhang CM, Zhang M, Nietfeldt J, Southward CM, Surette MG, Kachman SD, Benson AK. 2009. Genomic instability in regions adjacent to a highly conserved pch prophage in *Escherichia coli* O157:H7 generates diversity in expression patterns of the LEE pathogenicity island. *J. Bacteriol.* 191:3553–3568.
49. Leotta GA, Miliwsky ES, Chinen I, Espinosa EM, Azzopardi K, Tennant SM, Robins-Browne RM, Rivas M. 2008. Characterisation of Shiga toxin-producing *Escherichia coli* O157 strains isolated from humans in Argentina, Australia and New Zealand. *BMC Microbiol.* 8:46. doi:10.1186/1471-2180-8-46.
50. Franz E, van Hoek AH, van der Wal FJ, de Boer A, Zwartkruis-Nahuis A, van der Zwaluw K, Aarts HJ, Heuvelink AE. 2012. Genetic features differentiating bovine, food, and human isolates of Shiga toxin-producing *Escherichia coli* O157 in The Netherlands. *J. Clin. Microbiol.* 50:772–780.
51. Yokoyama E, Hirai S, Hashimoto R, Uchimura M. 2012. Clade analysis of enterohemorrhagic *Escherichia coli* serotype O157:H7/H- strains and hierarchy of their phylogenetic relationships. *Infect. Genet. Evol.* 12:1724–1728.
52. Vanaja SK, Springman AC, Besser TE, Whittam TS, Manning SD. 2010. Differential expression of virulence and stress fitness genes between *Escherichia coli* O157:H7 strains with clinical or bovine-biased genotypes. *Appl. Environ. Microbiol.* 76:60–68.
53. Fuller CA, Pellino CA, Flagler MJ, Strasser JE, Weiss AA. 2011. Shiga toxin subtypes display dramatic differences in potency. *Infect. Immun.* 79:1329–1337.
54. Besser TE, Shaikh N, Holt NJ, Tarr PI, Konkel ME, Malik-Kale P, Walsh CW, Whittam TS, Bono JL. 2007. Greater diversity of Shiga toxin-encoding bacteriophage insertion sites among *Escherichia coli* O157:H7 isolates from cattle than in those from humans. *Appl. Environ. Microbiol.* 73:671–679.
55. Mellor GE, Duffy LL, Dykes GA, Fegan N. 2010. Relative prevalence of *Salmonella* Sofia on broiler chickens pre- and postprocessing in Australia. *Poult. Sci.* 89:1544–1548.