JCM
Journals.ASM.org

# Geographically Distinct *Escherichia coli* O157 Isolates Differ by Lineage, Shiga Toxin Genotype, and Total Shiga Toxin Production

**Glen E. Mellor,[a] Narelle Fegan,[b] Kari S. Gobius,[b] Helen V. Smith,[c] Amy V. Jennison,[c] Beatriz A. D'Astek,[d] Marta Rivas,[d] Smriti Shringi,[e] Katherine N. K. Baker,[e] Thomas E. Besser[e]**

CSIRO Food and Nutrition Flagship, Archerfield, QLD, Australia[a]; CSIRO Food and Nutrition Flagship, Werribee, VIC, Australia[b]; Forensic and Scientific Services, Department of Health, Archerfield, QLD, Australia[c]; Servicio Fisiopatogenia, Instituto Nacional de Enfermedades Infecciosas-ANLIS "Dr. Carlos G. Malbrán," Buenos Aires, Argentina[d]; Washington State University, Department of Veterinary Microbiology and Pathology, Pullman, Washington, USA[e]

**While the differential association of *Escherichia coli* O157 genotypes with animal and human hosts has recently been well documented, little is known about their distribution between countries and how this might affect regional disease rates. Here, we used a 48-plex single nucleotide polymorphism (SNP) assay to segregate 148 *E. coli* O157 isolates from Australia, Argentina, and the United States into 11 SNP lineages. We also investigated the relationship between SNP lineages, Shiga toxin (Stx) gene profiles, and total Stx production. *E. coli* O157 isolates clearly segregated into SNP lineages that were differentially associated with each country. Of the 11 SNP lineages, seven were detected among isolates from a single country, two were detected among isolates from all three countries, and another two were detected only among U.S. and Argentinean isolates. A number of Australian (30%) and Argentinean (14%) isolates were associated with novel, previously undescribed SNP lineages that were unique to each country. Isolates within SNP lineages that were strongly associated with the carriage of $stx_{2a}$ produced comparatively more Stx on average than did those lacking the $stx_{2a}$ subtype. Furthermore, the proportion of isolates in $stx_{2a}$-associated SNP lineages was significantly higher in Argentina and the United States than Australia ($P < 0.05$). This study provides evidence for the geographic divergence of *E. coli* O157 and for a prominent role of $stx_{2a}$ in total Stx production. These results also highlight the need for more comprehensive studies of the global distribution of *E. coli* O157 lineages and the impacts of regionally predominant *E. coli* O157 lineages on the prevalence and severity of disease.**

Cattle are known to harbor *Escherichia coli* O157:H7 genotypes that are commonly implicated in human disease (1, 2) and are often suspected, either directly or indirectly, of being vehicles for transmission to humans (3, 4). Isolates derived from cattle typically exhibit greater genotypic diversity than do their human counterparts, and while similar genotypes occur in the two sources, only a subset of those seen in cattle are frequently represented in the human population (1, 5). The identification of markers that occur at different frequencies in cattle and human isolates represents an established strategy for identifying *E. coli* O157 genotypes of increased public health risk. In recent years, phylogenetics has proven particularly useful in identifying host-specific (cattle/human) associations among *E. coli* O157 populations (1, 5–10).

Over the past decade, numerous studies have reported associations between lineages or clades and the host source of *E. coli* O157. These studies have progressively become more complex and comprehensive as they aim to identify increasingly meaningful host associations. In an early phylogenetic study, a lineage-specific polymorphism assay (LSPA-6) (9, 11) that targeted six variably sized alleles within the *E. coli* O157 genome was developed to segregate *E. coli* O157 into two lineages; lineage I (LI) and lineage II (LII). Comparative genomic hybridization was later used to divide *E. coli* O157 isolates into three distinct clusters, which comprised lineages I, II, and I/II (12). The LSPA-6 assay has been used to report different cattle and human host associations among *E. coli* O157 populations from within the same country and between different countries. For instance, studies from the United States, Netherlands, and Japan have reported a dominance of LSPA-6 LII isolates in cattle populations and LI or LI/II in human populations (11, 13, 14). In contrast, LI/II was shown to be the dominant lineage among Australian and Argentinean cattle and human populations (15, 16). Similar studies from the United States and Canada have reported a dominance of LSPA-6 LI isolates in both cattle and human populations, while LII isolates were mostly associated with cattle sources (6, 9, 17). More recently, Manning et al. used an array of single nucleotide polymorphisms (SNPs) to categorize clinical isolates into nine distinct clades, of which one (clade 8) was significantly associated with hemolytic-uremic syndrome (HUS) in humans (18). The authors suggested that clade 8 isolates may have acquired characteristics that facilitate enhanced virulence in humans. In a subsequent study, Bono et al. (1) developed a panel of SNPs for the genotypic characterization of cattle- and human-derived *E. coli* O157 isolates. This was then used to categorize 426 isolates into eight major lineages which were grouped according to polymorphism-derived genotypes (1). Of the eight lineages, two were regularly represented in

**Received** 11 June 2014 **Returned for modification** 12 July 2014
**Accepted** 5 December 2014

**Accepted manuscript posted online** 10 December 2014

**Citation** Mellor GE, Fegan N, Gobius KS, Smith HV, Jennison AV, D'Astek BA, Rivas M, Shringi S, Baker KNK, Besser TE. 2015. Geographically distinct *Escherichia coli* O157 isolates differ by lineage, Shiga toxin genotype, and total Shiga toxin production. J Clin Microbiol 53:579–586. doi:10.1128/JCM.01532-14.

**Editor:** D. J. Diekema

Address correspondence to Glen E. Mellor, glen.mellor@csiro.au.

Supplemental material for this article may be found at http://dx.doi.org/10.1128/JCM.01532-14.



doi:10.1128/JCM.01532-14

humans and cattle and one was frequently represented in cattle but rarely associated with humans. Jung et al. (10) later expanded on this assay by developing a 48-locus SNP panel to incorporate additional genotypes (5) that were not originally represented by Bono et al. (1). This modified approach was subsequently used to group 530 *E. coli* O157 isolates into 11 SNP lineages (10), several of which were associated with a nonrandom distribution of cattle and human isolates.

The carriage and expression of virulence genes that encode Shiga toxins (Stx) by *E. coli* O157:H7 are essential for the development of bloody diarrhea as well as the severe and life-threatening disease in humans commonly referred to as hemolytic-uremic syndrome (HUS). Although essential to the development of HUS, most Shiga-toxigenic *E. coli* O157 infections do not progress to HUS, and the pathogenic potential of an isolate has been linked to the carriage of specific *stx* subtypes (19, 20), the quantity of Stx that they produce (21), and the lineage to which they belong (1, 10, 18). Shiga toxins, encoded by genes located in lambdoid bacteriophages, consist of two major types, Stx1 and Stx2, each of which includes multiple subtypes. Early epidemiological studies suggested that Shiga-toxigenic *E. coli* isolates carrying $stx_2$ were more frequently associated with severe disease outcomes in humans, such as HUS and bloody diarrhea, than those carrying $stx_1$ (22). Although $stx_2$ is common in *E. coli* O157 isolates from clinically ill humans, patients can present with a broad range of symptoms. The severity of disease caused by isolates may be attributed to the $stx_2$ variant(s) carried by *E. coli* O157 (23). Of the many toxin variants represented by the $stx_2$ branch, two subtypes, $stx_{2a}$ and $stx_{2c}$, predominate in cattle and human *E. coli* O157 isolates from across the globe. In a recent study by Jung et al. (10), the carriage of $stx_{2a}$ and $stx_{2c}$ was overrepresented in isolates of human- and cattle-associated SNP lineages, respectively. Variation in the toxicity of Stx types (24) and subtypes (25) has also been reported, with Stx2 being more potent than Stx1 and Stx2a demonstrating elevated potency compared to that of Stx2c (25). Furthermore, isolates representing human-biased genotypes, which typically carry $stx_{2a}$, are reported to produce significantly more total Stx than cattle-biased genotypes (21). In addition to host source association, differences in isolates from geographically separate countries have also been found, particularly those from Australia and the United States (5, 11, 15, 16).

Understanding the distribution of clinically associated genotypes from geographically separated countries, particularly those with differing incidences of disease, may provide a basis for the different disease burdens observed globally. In this study, we have used the assay developed by Jung et al. (10) to determine the distribution of SNP lineages in cattle and clinical isolates from Argentina, Australia, and the United States. We have also evaluated the relationship between *stx* subtypes and total Shiga toxin production. We hypothesize that genotypic differences associated with isolates of differing host and geographical sources, in addition to the presence and production of associated *stx* subtypes, may provide a basis for differences in the incidence and severity of *E. coli* O157-associated disease in each region.

## MATERIALS AND METHODS

**Isolate selection.** In total, 148 *E. coli* O157 isolates (see Table S1 in the supplemental material) representing similar numbers of Australian, Argentinean, and U.S. cattle and human isolates were chosen at random using Stata 12.1 (Stata Corporation, College Station, TX, USA) from previously well-characterized strain sets (15, 16).

**Single nucleotide polymorphism typing.** SNPs analyzed in this study were drawn from those discovered and characterized previously, as follows. Pooled DNA from 30 to 50 *E. coli* O157:H7 isolates per pool was sequenced, and reads were assembled against the Sakai strain reference genome to identify SNPs (1, 10). The phylogeny of the strains represented in the pooled DNA was established by determining the SNP haplotypes of large numbers of additional isolates (1). A reduced panel of SNPs capable of identifying the most prevalent lineage branches within the larger phylogeny was then selected and implemented in a custom designed 48-SNP GoldenGate assay (Illumina, San Diego, CA) (10). Genomic DNA was extracted from isolates using the Wizard genomic DNA purification kit (Promega, Alexandria, Australia), and 10-μl aliquots of purified DNA (50 ng/μl) were supplied to the University of Idaho iBEST Genomics Resources Core (Moscow, ID, USA) for analysis (BeadArray reader and the Genome Studio genotyping module v1.0; Illumina). SNP lineages were assigned Roman numerals corresponding to those designated by Bono et al. (1), while the subgroups within each SNP lineage were identified alphabetically as previously described (10) (Table 1). Isolates within groups that did not conform to a single SNP lineage or a subgroup within a SNP lineage were considered variants and represented by placing "var" after the SNP lineage to which they were most closely related.

**Shiga toxin gene subtyping.** Shiga toxin gene subtyping results from previous studies (15, 16) were used to examine relationships between SNP lineages and total Stx production.

**Shiga toxin production.** Total Stx production was assessed using a method adapted from the work of Shringi et al. (21). Specifically, overnight cultures were prepared with a single colony in 5 ml sterile Luria-Bertani Miller broth (LB broth; Oxoid, Basingstoke, United Kingdom) in 10-ml centrifuge tubes (Sarstedt, SA, Australia). Cultures were incubated at 37°C with rotary shaking (250 rpm) for 18 ± 2 h, after which 500 μl was transferred into 10 ml of sterile LB broth, mixed thoroughly, and processed for bacterial counts to ensure that the starting counts prior to Stx induction were not significantly different. A 1-ml aliquot of each subculture was immediately transferred into single wells of a 2-ml square deep-well plate (Sarstedt). Cultures were then induced with mitomycin C (Sigma-Aldrich Corp., St. Louis, MO, USA) at a final concentration of 0.5 μg $ml^{-1}$ for 24 ± 2 h at 37°C with rotary shaking (250 rpm). An unperforated plate sealer (Nunc, Roskilde, Denmark) was applied to the top of each deep-well plate prior to incubation to prevent evaporation. Following induction, cells were lysed with polymyxin B sulfate at 0.5 mg/ml to release periplasmic toxin as previously described (7, 26). The plate sealer was replaced, and cultures were incubated at 37°C for 1 h with rotary shaking (250 rpm). Polymyxin B-treated cultures were diluted 1:100 in sterile LB broth, immediately followed by a 1:2 dilution in sample diluent (Premier EHEC; Meridian Bioscience Inc., Cincinnati, Ohio). To confirm the linearity of the assay, serial dilutions were prepared from induced cultures of isolates containing $stx_1$ alone (strain E10705), $stx_{2a}$ alone (strain E2325), and $stx_{2c}$ alone (strain E3668). Total Stx production was measured using a Premier EHEC enzyme-linked immunosorbent assay (ELISA) (Meridian Bioscience Inc.) according to the manufacturer's instructions. Absorbance readings were obtained at dual wavelengths (450 nm/630 nm) using a Victor X microtiter plate reader (PerkinElmer, Glen Waverley, Australia). The mean values of two independent replicates from each isolate were used in subsequent analyses.

The presence or absence of separate Stx1 and Stx2 toxin was then assessed in mitomycin C-induced cultures using a commercially available Immuno*Card* STAT! EHEC test (Meridian). Briefly, a 175-μl volume of the polymyxin B-treated sample described above was diluted in 150 μl of sample diluent provided in the product kit. A 175-μl volume of each subculture was then transferred to the sample port on the test card and incubated at room temperature for 20 min. The presence of Stx1 and Stx2 was visually determined according to the manufacturer's instructions.

**TABLE 1** Distribution of SNP lineages in Australian, Argentinean, and U.S. cattle and human *E. coli* O157 isolates

| SNP lineage[b] | *n* | Previously identified host association[c] | No. (%) of isolates from country and host[a] | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | Australia | | United States | | Argentina | | Countries combined | |
| | | | Cattle | Human | Cattle | Human | Cattle | Human | Cattle | Human |
| Va | 3 | Unclassified[d] | 3 (13) | 0 | 0 | 0 | 0 | 0 | 3 (4) | 0 |
| Vb | 10 | Cattle | 0 | 0 | **9 (36)** | **1 (4)** | 0 | 0 | **9 (12)** | **1 (1)** |
| IVb | 24 | Unclassified | **5 (20)** | **19 (76)** | 0 | 0 | 0 | 0 | **5 (7)** | **19 (25)** |
| IVc | 2 | Unclassified | 0 | 0 | 1 (4) | 1 (4) | 0 | 0 | 1 (1) | 1 (1) |
| IVc var | 15 | Not detected[e] | **11 (44)** | **4 (16)** | 0 | 0 | 0 | 0 | 11 (15) | 4 (5) |
| IIb | 20 | Human | 0 | 0 | 0 | 1 (4) | 6 (24) | 13 (52) | 6 (8) | 14 (19) |
| IIa | 13 | Human | 0 | 0 | 0 | 1 (4) | 4 (16) | 8 (32) | 4 (5) | 9 (12) |
| IIa var | 5 | Not detected | 0 | 0 | 0 | 0 | 4 (16) | 1 (4) | 4 (5) | 1 (1) |
| Ia | 16 | Cattle | 4 (16) | 1 (4) | 1 (4) | 0 | 8 (32) | 2 (8) | **13 (18)** | **3 (4)** |
| Ib | 38 | Human | 1 (4) | 1 (4) | **13 (52)** | **21 (84)** | 1 (4) | 1 (4) | 15 (21) | 23 (31) |
| Ib var | 2 | Not detected | 0 | 0 | 0 | 0 | 2 (8) | 0 | 2 (3) | 0 |
| Total | 148 | NA[f] | 24 (100) | 25 (100) | 24 (100) | 25 (100) | 25 (100) | 25 (100) | 73 (100) | 75 (100) |

[a] Bold text indicates significant differences in the distribution of SNP lineages between cattle and human isolates within each country or between the combined totals of cattle and human isolates.
[b] With the exception of SNP lineages IVc var, IIa var, and Ib var, all other SNP lineages refer to those which were previously defined by Jung et al. (10).
[c] SNP lineage associations with hosts that have previously been reported by Jung et al. (10).
[d] SNP lineages were sparsely represented in the original study by Jung et al. (10) and were subsequently unable to be tested for association with a specific host.
[e] SNP lineages were not detected in any of the isolates previously reported by Jung et al. (10).
[f] NA, not applicable.

**Statistical analysis.** The distribution of SNP lineages in host sources, countries, or *stx* genotypes was evaluated using chi-squared tests ($\chi^2$) on 2-by-2 or 2-by-3 contingency tables (Stata 12.1; Stata Corporation, College Station, TX, USA). Only SNP lineages containing at least 10 isolates (Ia, Ib, IIa, IIb, IVb, IVc var, and Vb) were used to evaluate SNP lineage associations with host sources or countries. *P* values were two-tailed, and groups were considered significantly different if *P* values were $<0.05$.

## RESULTS

**Distribution of lineages and subclades.** All 148 *E. coli* O157 isolates were typed using a panel of 48 SNPs to segregate isolates into 11 SNP lineages (Fig. 1). Of the 148 isolates examined, 22 (15%) carried allele profiles intermediate between two previously characterized SNP lineages; these isolates were designated Ib variant, IIa variant, and IVc variant lineages based on their position in the phylogenetic tree. SNP lineage Ib variant isolates possessed only one (JB2419218) of the four SNP locus alleles that differentiate lineage Ib from Ia (10). Lineage IIa variant isolates possessed two of the three SNP locus alleles expected of lineage IIa, but the third SNP locus (JB1107987) was that expected of the immediate ancestor (10). Finally, the SNP lineage IVc variant possessed only one (TEB869577) of the two SNP locus alleles that differentiate lineage IVc from lineage IVb (10).

Isolates from Australia, Argentina, and the United States were comprised of five, six, and six SNP lineages, respectively (Fig. 1). Of the 11 SNP lineages detected, seven were detected only in a single country, two were detected in the United States and Argentina, and two were detected across all three countries. All SNP lineages containing 10 or more isolates (i.e., Ia, Ib, IIa, IIb, IVb, IVc var, and Vb) demonstrated significant associations with countries (Fig. 1). The majority of isolates within each country were associated with different, predominant SNP lineages. SNP lineages IVb (48%) and IVc var (30%) were frequent among Australian isolates but absent from the Argentinean and U.S. isolate sets. All three SNP lineages that accounted for the majority of Argentinean isolates, IIb (38%), IIa (24%), and Ia (20%), were also present in the United States but were less frequent (single isolate only), while SNP lineage Ia was present in only 10% of Australian isolates. Of the two SNP lineages frequently detected in the United States (Ib, 68%, and Vb, 20%), isolates belonging to Ib were also detected in Australia (4%) and Argentina (4%) but at much lower frequencies.

Significant associations between SNP lineages and host sources (cattle and human) were identified in Australia and the United States but not Argentina (Table 1). Specifically, Australian isolates belonging to SNP lineage IVb were overrepresented in humans ($P < 0.05$) while those belonging to IVc var were overrepresented in cattle ($P < 0.05$). Likewise, U.S. isolates belonging to SNP lineages Ib and Vb were overrepresented in humans ($P < 0.05$) and cattle ($P < 0.05$), respectively. While the findings were not statistically significant, more Argentinean human isolates were detected in SNP lineages IIa (32% human versus 16% cattle) and IIb (52% human versus 24% cattle) while cattle isolates were more frequent in Ia (32% versus 8%). When cattle and human isolates were combined from all three countries, only SNP lineages IVb and Vb retained their respective human- and cattle-host associations. In addition, a new host association identified as SNP lineage Ia was significantly associated with cattle ($P < 0.05$).

**Distribution of *stx* genes within SNP lineages.** The majority of isolates (110 of 148, 74%) were shown to carry two or more different Shiga toxin genes, and the prevalences of $stx_{2c}$ (98 of 148, 66%), $stx_1$ (88 of 148, 59%), and $stx_{2a}$ (75 of 148, 51%) were similar among the isolates tested. Despite this, a nonrandom distribution of *stx* subtypes was detected across the SNP lineages. Of the seven SNP lineages containing 10 or more isolates, four were significantly associated ($P < 0.05$) with the carriage of $stx_{2c}$ (Ia, IVb, IVc var, and Vb) either in the presence or in the absence of $stx_1$, while three were significantly associated ($P < 0.05$) with the carriage of $stx_{2a}$ (Ib, IIa, and IIb) either in the presence or in the absence of $stx_1$ and $stx_{2c}$ (Fig. 2). Greater than 80% of isolates

| | No. (%) of *E. coli* O157 | | |
|---|---|---|---|
| | **Australia** | **U.S.** | **Argentina** |
| VII (III, VIII) | 0 | 0 | 0 |
| 1 VI | 0 | 0 | 0 |
| 2 Va | 3 (8) | 0 | 0 |
| 3 Vb | 0 | 10 (20) | 0 |
| 4 IVa | 0 | 0 | 0 |
| 5 IVb | 24 (48) | 0 | 0 |
| 6 IVc | 0 | 2(4) | 0 |
| **IVc variant** | 15 (30) | 0 | 0 |
| 7 IIb | 0 | 1 (2) | 19 (38) |
| 8 9 IIa | 0 | 1 (2) | 12 (24) |
| 10 **IIa variant** | 0 | 0 | 5 (10) |
| 11 Ia | 5 (10) | 1 (2) | 10 (20) |
| 12 Ib | 2 (4) | 34 (68) | 2 (4) |
| **Ib variant** | 0 | 0 | 2 (4) |

**FIG 1** Schematic illustration of the phylogenetic relationship of *E. coli* O157 isolates derived from Australia, the United States, and Argentina based on SNP-defined lineages. The numbers (percentages) of isolates from each country that belong to the various SNP lineages are shown below country headings. The diagram was adapted from the work of Jung et al. (10) to include the three variant SNP lineages identified in the current study (Ib variant, IIa variant, and IVc variant), whose branch points from the previously characterized lineages were determined by their SNP allele composition. The numbers on the branches reference the specific SNP loci that support each bifurcation (see Table S2 in reference 10); the SNP alleles that characterized the variant lineages identified here are JB2419218, JB1107987, and TEB869577, as explained in the Results. SNP lineages that contained more than 10 isolates were tested for significance. Those that were significantly associated with a country are represented by boxes.

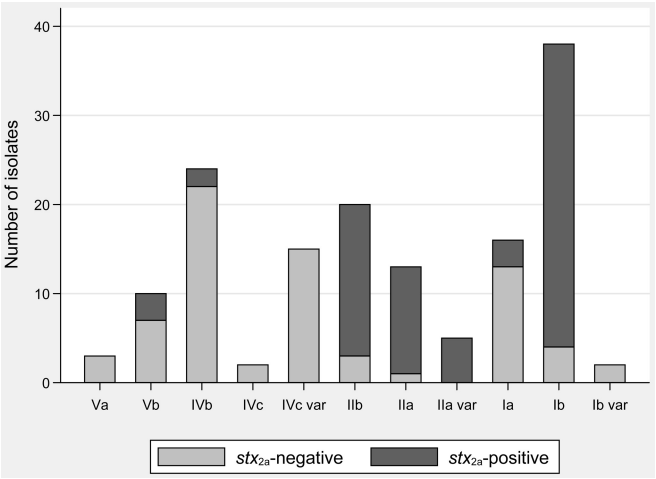


**FIG 2** Number of *E. coli* O157 isolates within each SNP lineage that tested positive or negative by PCR for the presence of the $stx_{2a}$ gene. A total of 148 strains were screened for *stx* genes, of which 72 were $stx_{2a}$ negative and 76 were $stx_{2a}$ positive. $stx_{2a}$-negative genotypes included $stx_1$ alone, $stx_1$ and $stx_{2c}$, or $stx_{2c}$ alone. $stx_{2a}$-positive genotypes included isolates that possessed $stx_{2a}$ alone or in combination with $stx_1$, $stx_{2c}$, or both. Individual isolate genotype data are available in Table S1 in the supplemental material.

within SNP lineages IVb (20 of 24, 83%) and IVc var (14 of 15, 93%) carried both $stx_1$ and $stx_{2c}$. Similarly, isolates within SNP lineage Ib were strongly associated with the carriage of both $stx_1$ and $stx_{2a}$ (30 of 38, 79%) while those belonging to SNP lineages IIa (10 of 13, 77%) and IIb (13 of 20, 65%) were both strongly associated with the carriage of both $stx_{2a}$ and $stx_{2c}$ subtypes.

**Total Shiga toxin production.** Commercially available immunoassays were used to determine the presence of Stx1 and Stx2 and measure the relative Stx production of all isolates included in the study. The Premier EHEC assay was shown to be linear for Stx1 ($R^2 = 0.9997$), Stx2a ($R^2 = 0.9897$), and Stx2c ($R^2 = 0.9734$) when values were in the range of ~0.5 and ~2.5 absorbance units. The majority of isolates carrying both $stx_1$ and $stx_2$ (including $stx_{2a}$ and $stx_{2c}$ subtypes) tested positive for the production of both Stx types. However, the mean production of Stx in isolates that possessed $stx_{2a}$ was comparatively greater than the production in those that lacked the $stx_{2a}$ subtype. This trend was also apparent across individual *stx* genotypes, with the mean Stx production of isolates belonging to $stx_{2a}$-positive genotypes ($stx_{2a}$, $stx_{2a}$ $stx_{2c}$, $stx_1$ $stx_{2a}$, and $stx_1$ $stx_{2a}$ $stx_{2c}$) comparatively higher than the mean production of each $stx_{2a}$-negative genotype ($stx_1$, $stx_1$ $stx_{2c}$, and $stx_{2c}$) (Fig. 3A). In total, 74 of the 76 (97%) $stx_{2a}$-positive isolates exhibited high levels of Stx approaching the saturation threshold (~2.5 absorbance units) of the test system used in this study and therefore

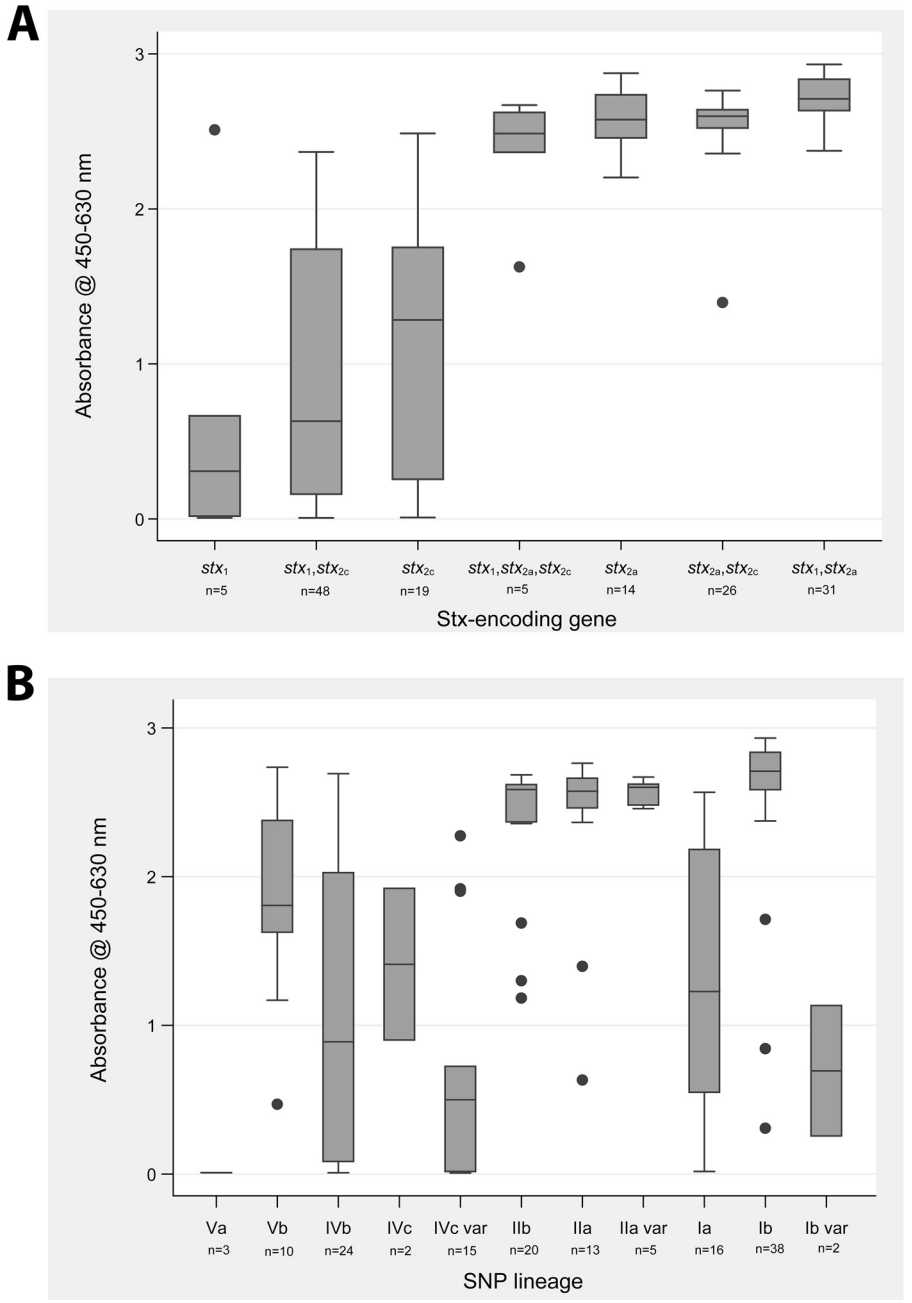


**FIG 3** Box plots displaying the total Stx production of isolates within Shiga toxin genotypes (A) and SNP lineages (B). Total Stx production was measured in duplicate for each isolate using the Premier EHEC ELISA. Within each box, the inside line represents the median and the lower and upper hinges of each box represent the 25th (Q1) and 75th (Q3) percentiles, respectively. The interquartile range (IQR) is represented by Q3 − Q1. The upper and lower adjacent lines (whiskers) represent Q3 + (1.5 × IQR) and Q1 − (1.5 × IQR), respectively. Outliers are represented by solid circles if data points are above Q3 + (1.5 × IQR) or below Q1 − (1.5 × IQR).

limiting the variance of the readings. Despite having a lower mean value than $stx_{2a}$-positive genotypes, isolates belonging to all three $stx_{2a}$-negative genotypes exhibited absorbance values of <0.15 to greater than 2.0.

Stx production also varied widely among isolates within each SNP lineage. Isolates belonging to SNP lineages that were significantly associated with the carriage of $stx_{2a}$ (Ib, IIa, and IIb) had the highest mean production of Stx (Fig. 3B). Overall, 92% of isolates in Ib, 85% of isolates in IIa, and 85% of isolates in IIb produced high levels (>2.0 absorbance units) of Stx. However, all three SNP lineages also contained a small proportion of isolates representing the lower extremes of Stx production. Isolates belonging to SNP lineages Ia and Vb displayed mean absorbance values in the range of 1.0 to 2.0 absorbance units, while those belonging to IVb and IVc var produced relatively low mean quantities of Stx (<1.0 absorbance units) compared to most other SNP lineages. Four SNP lineages (Ib var, IIa var, IVc, and Va) contained too few isolates to make meaningful comparisons.

## DISCUSSION

In this study, we demonstrate that *E. coli* O157 isolates obtained from both cattle and humans in different geographic locations (Australia versus United States versus Argentina) largely represent different SNP lineages. Since isolates used here represent a subset of those previously shown to separate by various typing methods (15, 16), it was not surprising to achieve a similar level of SNP lineage segregation in the current study. However, we have now further defined additional specific genomic markers, correlated with geographic segregation, by the identification of three novel SNP haplotypes that identify the branch points of novel lineages (Ib var, IIa var, and IVc var); it is likely that whole-genome sequencing of isolates of these lineages would reveal novel lineage-associated SNPs. While variant SNP lineages Ib var and IIa var collectively accounted for a small proportion of Argentinean isolates (14%), IVc var was the second most prevalent SNP lineage identified in Australian isolates (30%). In a previous study, isolates belonging to Ib var, IIa var, and IVc var were not detected in any of the 530 isolates screened (10), indicating that these variants may be exclusive to either Argentina (Ib var and IIa var) or Australia (IVc var). Pronounced differences in the distribution of SNP lineages among Australian, Canadian, New Zealand, and Japanese isolates were also identified in the study by Jung et al. (10). The authors also reported the observation of a variant Japanese SNP lineage (5/31 isolates, 16%) intermediate between lineages Va and Vb, with one allele expected of the former and two alleles expected of the latter. In addition to this, evidence for further geographic specialization has also been reported in Australian isolates. The integration of $stx_1$ phage in a locus that has previously been associated only with $stx_{2a}$ phage and the variability of *q* promoter regions upstream of $stx_{2c}$ appear to be unique to Australian *E. coli* O157 isolates (15, 16). The occurrence of similar genotypes at different frequencies in isolates obtained from different geographic locations points to the early global dissemination of *E. coli* O157. However, the subsequent emergence of predominant and sometimes exclusive genotypes in several countries suggests that isolates may have responded to regional selective pressure to form divergent evolutionary populations, though it is equally possible that geographic divergence may be a product of genetic drift or founder effects.

Of the 11 SNP lineages identified by Jung et al. (10), two (Ia and Vb) were overrepresented in cattle isolates and three (Ib, IIa, and IIb) were overrepresented in human isolates. A similar nonrandom distribution of cattle and human isolates across SNP lineages was also confirmed in the current study. Nevertheless, a notable exception was observed in Argentina, where the distribution of lineages IIa and IIb was not significantly different between cattle and humans. Although statistical significance was not reached for these lineages, the host association tendencies reported here (association of SNP lineages IIa and IIb with human clinical isolates) are the same as those reported by Jung et al. (10).

Strong correlations were also observed between SNP lineages and country of origin. Compared to Australia and the United States, Argentina had a relatively high prevalence of *E. coli* O157 isolates belonging to human-associated SNP lineages IIa and IIb. Using a similar SNP typing scheme, isolates belonging to these two SNP lineages were previously grouped together into a single hypervirulent clade (clade 8) (10, 16). The hypervirulent classification of clade 8 isolates is due to their significant association with HUS in humans (18) and may also be linked to increased Stx production (27). Since Argentina has the highest reported incidence of HUS in children at or below the age of 5 years (28), our findings indirectly support the link between HUS and clade 8 (SNP lineages IIa and IIb) observed by Manning et al. (18). As beef cattle are considered a major source of *E. coli* O157 transmission to humans, we hypothesize that the high frequency of lineages IIa and IIb in Argentinean cattle would result in a predominance of human exposure to pathogenic lineages, a result of which would influence the incidence of HUS in Argentina. Similarly, the majority of U.S. cattle and human isolates were shown to belong to SNP lineage Ib, which is frequently associated with clinical disease in the United States (10). In comparison, Australia has a relatively low reported frequency of HUS cases (29) and the incidence of disease associated with *E. coli* O157 appears to be lower than the global average. Although the predominant SNP lineages identified in Australia have not previously been associated with human sources (10), a significant association of SNP lineage IVb with Australian humans was identified in the current study. However, in contrast to the study by Jung et al. (10), which comprised 530 isolates, the current study investigated a smaller number of *E. coli* O157 strains and further investigation into the distribution and host associations of SNP lineages in Australia is required to validate this finding.

In the current study, isolates belonging to previously identified human (Ib, IIa, and IIb)- and cattle (Ia and Vb)-associated SNP lineages were also significantly associated with the carriage of $stx_{2a}$ (89%) or $stx_{2c}$ (77%), respectively. These results are consistent with those reported previously (1, 10) and provide indications of *stx* stability within genomically divergent populations. Since isolates carrying $stx_{2c}$ are known to be less pathogenic than those carrying $stx_{2a}$ (19, 20), it is plausible that the increased occurrence of human disease caused by isolates within SNP lineages Ib, IIa, and IIb may be attributed to the Stx complement(s) of these lineages (1, 10) and/or the quantity of Stx that they produce (21, 27). Here, *E. coli* O157 isolates belonging to $stx_{2a}$-positive genotypes were shown to produce comparatively more total Stx on average than those belonging to $stx_{2a}$-negative genotypes. Isolates carrying the $stx_{2a}$ subtype consistently (97.3%) produced high levels of Stx, regardless of the presence of additional *stx* subtypes, their SNP lineage designation, or their country of origin. Since $stx_{2a}$ is highly associated with clinical SNP lineages, it is not unexpected that these isolates also produced the highest mean quantities of Stx. In contrast, isolates belonging to cattle-associated SNP lineages (Ia and Vb) produced lower mean quantities of Stx. Australian isolates were predominantly associated with SNP lineages containing $stx_{2c}$, and the mean production of Stx from this country was lower than that observed in Argentina and the United States, where $stx_{2a}$ was detected in a higher proportion of isolates. These findings are consistent with those of Shringi et al. (21), who demonstrated higher total Stx production in *E. coli* O157 associated with clinical disease and hence $stx_{2a}$. Isolates belonging to human-associated SNP lineages frequently possess the $stx_{2a}$ gene, and its protein complement is reported to be comparatively more toxic than Stx1 and Stx2c (25). Therefore, it is feasible that the dual effects of two factors, the presence of $stx_{2a}$ in isolates with high Stx production and the increased toxicity of Stx2a, act to enhance the virulence of isolates within human-associated SNP lineages. However, it should be noted that no attempt was made to determine if differences in absorbance values were due to varying sensitivity of the

assay to different toxin variants (Stx1, Stx2a, and Stx2c). Since we demonstrate that the frequency of the toxin variants differed between countries and SNP lineages, differences in the sensitivities of Stx variants are a possible alternative explanation for the variance in absorbance readings observed across the various groups examined.

An additional two SNP lineages were shown to demonstrate cattle (IVc var) and human (IVb) host associations in the current study that were not identified previously (10). However, in contrast to the previously identified host-associated lineages, these additional lineages did not predominantly possess different *stx* subtypes or produce vastly different quantities of Stx. For example, Australian human isolates were significantly associated with SNP lineage IVb and predominantly possessed *stx* combinations ($stx_1$ and $stx_{2c}$ but not $stx_{2a}$) more consistent with cattle than human-associated genotypes. The predominant *stx* genotype associated with this lineage is also identical to that of isolates belonging to cattle-biased SNP lineage IVc var. Furthermore, isolates within SNP lineages IVb and IVc var are reported to produce similar mean quantities of Stx despite displaying different associations with host sources. In a previous study of mostly U.S. isolates, these SNP lineages were either not detected or present in numbers too low to be robustly tested for host associations (10). Although these SNP lineages are common to Australia, a larger set of isolates would need to be tested to confirm these host associations. Nonetheless, we believe that isolates belonging to SNP lineage IVb are likely to result in less frequent and severe disease outcomes (as determined by the frequency of HUS cases) in Australia based on *E. coli* O157 epidemiological data for this country (29). While compelling data exist to suggest that the pathogenic potential of *E. coli* O157 is related to the carriage of specific *stx* variants, additional underlying factors are also likely to influence lineage-associated virulence.

Currently, *E. coli* O157:H7 isolates are considered adulterants of raw nonintact beef products and product components in the United States. Recently, an additional six serotypes have been given similar adulterant status in the United States on the basis that they belong to a clinically significant serotype and that they possess the key virulence genes *stx* and *eae*. The data presented here and elsewhere suggest that *E. coli* O157 isolates differ in their association with humans despite possessing *stx* and *eae*. Therefore, these virulence markers may not be the best indicator of an isolate's pathogenic potential. As our understanding of *E. coli* pathogenicity advances, more definitive traits may be used to identify high-risk isolates that are likely to cause disease in humans. This may also lead to more-targeted regulation of these organisms in the beef cattle industry based on the likelihood of harm that different genotypes present to public health.

Together with our previous findings (15, 16), SNP data reported here from Australian, Argentinean, and U.S. isolates provide compelling evidence for the geographical segregation of *E. coli* O157. The occurrence of different predominant and sometimes exclusive phylotypes in each region also supports the global dissemination and subsequent divergent evolution of this organism in each country. Furthermore, the regional dominance of host-related SNP lineages may also provide a possible explanation for the lower disease burden seen in Australia (29) relative to the United States and Argentina (28, 30). This study not only provides a strong basis for the further exploration of regional variations in *E. coli* O157 genotypes and their associated impacts on the regional prevalence and severity of disease but also supports current genomic research in the global evolution and phylogeny of *E. coli* O157.

## ACKNOWLEDGMENTS

This study was supported by research funds from the Commonwealth Scientific and Industrial Research Organization (CSIRO), USDA AFRI grants 2009-04248 and 2010-04487, the Victorian Department of Environment and Primary Industries, and Meat and Livestock Australia.

## REFERENCES

1. **Bono JL, Smith TP, Keen JE, Harhay GP, McDaneld TG, Mandrell RE, Jung WK, Besser TE, Gerner-Smidt P, Bielaszewska M, Karch H, Clawson ML.** 2012. Phylogeny of Shiga toxin-producing *Escherichia coli* O157 isolated from cattle and clinically ill humans. Mol Biol Evol **29:** 2047–2062. http://dx.doi.org/10.1093/molbev/mss072.
2. **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. http://dx.doi.org/10.1128/AEM.01035-06.
3. **Ferens WA, Hovde CJ.** 2011. *Escherichia coli* O157:H7: animal reservoir and sources of human infection. Foodborne Pathog Dis **8:**465–487. http://dx.doi.org/10.1089/fpd.2010.0673.
4. **Garcia A, Fox JG, Besser TE.** 2010. Zoonotic enterohemorrhagic *Escherichia coli*: a One Health perspective. ILAR J **51:**221–232. http://dx.doi.org/10.1093/ilar.51.3.221.
5. **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. http://dx.doi.org/10.1128/AEM.01761-09.
6. **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. http://dx.doi.org/10.1128/AEM.02821-07.
7. **Zhang Y, Laing C, Zhang Z, Hallewell J, You C, Ziebell K, Johnson RP, Kropinski AM, Thomas JE, Karmali M, Gannon VPJ.** 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. http://dx.doi.org/10.1128/AEM.01288-09.
8. **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. http://dx.doi.org/10.1016/j.meegid.2010.10.007.
9. **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. http://dx.doi.org/10.1128/AEM.70.11.6846-6854.2004.
10. **Jung WK, Bono JL, Clawson ML, Leopold SR, Shringi S, Besser TE.** 2013. Lineage and genogroup-defining single nucleotide polymorphisms of *Escherichia coli* O157:H7. Appl Environ Microbiol **79:**7036–7041. http://dx.doi.org/10.1128/AEM.02173-13.
11. **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, β-glucuronidase-negative enterohemorrhagic *Escherichia coli* O157. J Bacteriol **183:**6885–6897. http://dx.doi.org/10.1128/JB.183.23.6885-6897.2001.
12. **Zhang Y, Laing C, Steele M, Ziebell K, Johnson R, Benson AK, Taboada E, Gannon VPJ.** 2007. Genome evolution in major *Escherichia coli* O157:H7 lineages. BMC Genomics **8:**121. http://dx.doi.org/10.1186/1471-2164-8-121.
13. **Lee K, French NP, Hara-Kudo Y, Iyoda S, Kobayashi H, Sugita-Konishi Y, Tsubone H, Kumagai S.** 2011. Multivariate analyses revealed distinctive features differentiating human and cattle isolates of Shiga toxin-producing *Escherichia coli* O157 in Japan. J Clin Microbiol **49:**1495–1500. http://dx.doi.org/10.1128/JCM.02640-10.
14. **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. http://dx.doi.org/10.1128/JCM.05964-11.

15. **Mellor GE, Besser TE, Davis MA, Beavis B, Jung W, Smith HV, Jennison AV, Doyle CJ, Chandry PS, Gobius KS, Fegan N.** 2013. Multilocus genotype analysis of *Escherichia coli* O157 isolates from Australia and the United States provides evidence of geographic divergence. Appl Environ Microbiol **79:**5050–5058. http://dx.doi.org/10.1128/AEM.01525-13.
16. **Mellor GE, Sim EM, Barlow RS, D'Astek 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. http://dx.doi.org/10.1128/AEM.00365-12.
17. **Sharma R, Stanford K, Louie M, Munns K, John SJ, Zhang Y, Gannon V, Chui L, Read R, Topp E, McAllister T.** 2009. *Escherichia coli* O157:H7 lineages in healthy beef and dairy cattle and clinical human cases in Alberta, Canada. J Food Prot **72:**601–607.
18. **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. http://dx.doi.org/10.1073/pnas.0710834105.
19. **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. http://dx.doi.org/10.1371/journal.pone.0051572.
20. **Friedrich AW, Bielaszewska M, Zhang WL, Pulz M, Kuczius T, Ammon A, Karch H.** 2002. *Escherichia coli* harboring Shiga toxin 2 gene variants: frequency and association with clinical symptoms. J Infect Dis **185:**74–84. http://dx.doi.org/10.1086/338115.
21. **Shringi S, Garcia 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. http://dx.doi.org/10.1128/IAI.05470-11.
22. **Boerlin P, McEwen SA, Boerlin-Petzold F, Wilson JB, Johnson RP, Gyles CL.** 1999. Associations between virulence factors of Shiga toxin-producing *Escherichia coli* and disease in humans. J Clin Microbiol **37:**497–503.
23. **Kawano K, Okada M, Haga T, Maeda K, Goto Y.** 2008. Relationship between pathogenicity for humans and stx genotype in Shiga toxin-producing *Escherichia coli* serotype O157. Eur J Clin Microbiol Infect Dis **27:**227–232. http://dx.doi.org/10.1007/s10096-007-0420-3.
24. **Tesh VL, Burris JA, Owens JW, Gordon VM, Wadolkowski EA, Obrien AD, Samuel JE.** 1993. Comparison of the relative toxicities of Shiga-like toxins type I and type II for mice. Infect Immun **61:**3392–3402.
25. **Fuller CA, Pellino CA, Flagler MJ, Strasser JE, Weiss AA.** 2011. Shiga toxin subtypes display dramatic differences in potency. Infect Immun **79:**1329–1337. http://dx.doi.org/10.1128/IAI.01182-10.
26. **Laing CR, Zhang YX, Gilmour MW, Allen V, Johnson R, Thomas JE, Gannon VPJ.** 2012. A comparison of Shiga-toxin 2 bacteriophage from classical enterohemorrhagic *Escherichia coli* serotypes and the German *E. coli* O104:H4 outbreak strain. PLoS One **7:**e37362. http://dx.doi.org/10.1371/journal.pone.0037362.
27. **Zangari T, Melton-Celsa AR, Panda A, Smith MA, Tatarov I, De Tolla L, O'Brien AD.** 2014. Enhanced virulence of the *Escherichia coli* O157:H7 spinach-associated outbreak strain in two animal models is associated with higher levels of Stx2 production after induction with ciprofloxacin. Infect Immun **82:**4968–4977. http://dx.doi.org/10.1128/IAI.02361-14.
28. **Rivas M, Chinen I, Miliwebsky E, Galli L, Repetto HA, Masana M.** 2011. Epidemiology of Argentinean STEC, p 109–132. *In* Walk S, Feng P (ed), Bacterial population genetics: a tribute to Thomas S. Whittam. ASM Press, Washington, DC.
29. **Vally H, Hall G, Dyda A, Raupach J, Knope K, Combs B, Desmarchelier P.** 2012. Epidemiology of Shiga toxin producing *Escherichia coli* in Australia, 2000–2010. BMC Public Health **12:**63. http://dx.doi.org/10.1186/1471-2458-12-63.
30. **Centers for Disease Control and 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.