Applied and Environmental Microbiology, Apr. 2008, p. 2153–2160 Vol. 74, No. 7
0099-2240/08/$08.00+0 doi:10.1128/AEM.02566-07

Molecular Analysis as an Aid To Assess the Public Health Risk of Non-O157 Shiga Toxin-Producing *Escherichia coli* Strains▽‡

Brian K. Coombes,[1,2]*† Mark E. Wickham,[3]†§ Mariola Mascarenhas,[2] Samantha Gruenheid,[4] B. Brett Finlay,[3] and Mohamed A. Karmali[2]

Department of Biochemistry and Biomedical Sciences and Michael G. DeGroote Institute for Infectious Disease Research, McMaster University, Hamilton, Ontario, Canada[1]; Laboratory for Foodborne Zoonoses, Public Health Agency of Canada, Guelph, Ontario, Canada[2]; Michael Smith Laboratories, University of British Columbia, Vancouver, British Columbia, Canada[3]; and Department of Microbiology and Immunology, McGill University, Montreal, Quebec, Canada[4]

Received 14 November 2007/Accepted 21 January 2008

Shiga toxin-producing *Escherichia coli* (STEC) strains are commensal bacteria in cattle with high potential for environmental and zoonotic transmission to humans. Although O157:H7 is the most common STEC serotype, there is growing concern over the emergence of more than 200 highly virulent non-O157 STEC serotypes that are globally distributed, several of which are associated with outbreaks and/or severe human illness such as hemolytic-uremic syndrome (HUS) and hemorrhagic colitis. At present, the underlying genetic basis of virulence potential in non-O157 STEC is unknown, although horizontal gene transfer and the acquisition of new pathogenicity islands are an expected origin. We used seropathotype classification as a framework to identify genetic elements that distinguish non-O157 STEC strains posing a serious risk to humans from STEC strains that are not associated with severe and epidemic disease. We report the identification of three genomic islands encoding non-LEE effector (*nle*) genes and 14 individual *nle* genes in non-O157 STEC strains that correlate independently with outbreak and HUS potential in humans. The implications for transmissible zoonotic spread and public health are discussed. These results and methods offer a molecular risk assessment strategy to rapidly recognize and respond to non-O157 STEC strains from environmental and animal sources that might pose serious public health risks to humans.

Shiga toxin-producing *Escherichia coli* (STEC) is a zoonotic food- and waterborne pathogen that is a serious public health concern because of its propensity to cause outbreaks, hemorrhagic colitis, and the potentially fatal hemolytic-uremic syndrome (HUS), the leading cause of renal failure in children (15, 25). The O157:H7 serotype of enterohemorrhagic STEC is the most common and is well recognized for its high virulence in human populations (25). Ruminants, especially cattle, are the main reservoir of STEC and can transmit the organism to humans by direct contact with infected animals or food products or via water contamination (15), as was the case in Walkerton, Ontario, Canada, when *E. coli* O157:H7 from livestock fecal material contaminated the local watershed (3). However, there is growing concern over the emergence of more than 200 non-O157 STEC serotypes associated with human illness (14, 30), occasionally severe or epidemic, in Canada, the United States, Australia, Latin America, and Europe (4, 5, 17, 33). At present, there is no capacity to differentiate high-risk non-O157 STEC serotypes from those posing little risk to humans.

To begin to assess the clinical and public health risks associated with non-O157 STEC, we previously developed seropathotype classification that considers serotype association with human epidemics, HUS, and diarrhea (16). Seropathotype A comprises serotypes O157:H7 and O157:NM, which are common causes of outbreaks and HUS in most countries; seropathotype B strains are associated with outbreaks and HUS, but less frequently than seropathotype A; seropathotype C strains are associated with sporadic HUS but not epidemics; seropathotype D strains are associated with diarrhea but not with outbreaks or HUS; and seropathotype E comprises multiple STEC serotypes that have never been associated with human disease and appear to be linked exclusively to animal infections in an agricultural setting.

While the production of Shiga toxin by STEC is the primary virulence trait responsible for HUS, many non-O157:H7 strains that produce Shiga toxin do not cause HUS, indicating additional virulence determinants are required for disease. Virulence in bacterial pathogens is modulated by acquisition of mobile genetic elements such as bacteriophages, transposons, plasmids, and genomic islands (20). One class of genomic islands, called pathogenicity islands (PAIs), harbor genes that are virulence factors during host infection (9, 12). PAIs constitute a flexible gene pool contributing to pathogen evolution and virulence potential and can be used as a genetic signature of new and emerging pathogens. In STEC, for example, the locus of enterocyte effacement (LEE) is a chromosomal PAI encoding a type III secretion system necessary for the attaching-and-effacing lesion that is pathognomonic of disease associated with this organism. Both O157:H7 (19, 36) and non-O157 STEC strains (16, 38) contain a variable repertoire of virulence determinants, including a collection of non-LEE-encoded effector (*nle*) genes that encode translocated substrates of the type III secretion system. However, the contri-

* Corresponding author. Mailing address: Department of Biochemistry and Biomedical Sciences, McMaster University, Health Sciences Centre 4H17, 1200 Main St. West, Hamilton, Ontario L8N 3Z5, Canada. Phone: (905) 525-9140, ext. 22159. Fax: (905) 522-9033. E-mail: coombes@mcmaster.ca.

† B.K.C. and M.E.W. contributed equally to this study.

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

§ Present address: Phillips, Ormonde, and Fitzpatrick, Melbourne, Australia.

▽ Published ahead of print on 1 February 2008.

TABLE 1. Oligonucleotides used for gene content analysis in non-O157 STEC

Gene target[a]	Oligonucleotide sequence (5′–3′)		Product size (bp)	Annealing temp (°C)
	Primer	Primer 2		
nleA	ATGAACATTCAACCGACCATAC	GACTCTTGTTTCTTGGATTATATCAAA	1,296	55
nleB	GGAAGTTTGTTTACAGAGACG	AAAATGCCGCTTGATACC	297	55
nleB2	GTTAATACTAAGCAGCATCC	CCATATCAAGATAGATACACC	475	52
nleC	ACAGTCCAACTTCAACTTTTCC	ATCGTACCCAGCCTTTCG	777	55
nleD	GGTATTACATCAGTCATCAAGG	TTGTGGAAAACATGGAGC	426	55
nleE	GTATAACCAGAGGAGTAGC	GATCTTACAACAAATGTCC	260	52
nleF	ATGTTACCAACAAGTGGTTCTTC	ATCCACATTGTAAAGATCCTTTGTT	567	55
nleG (Z6010)	ATGTTATCGCCCTCTTCTATAAAT	ACTTAATACTACACTAATAAGATCCA	902	55
nleG2-1 (Z6025)	ACCAGAAACCTGACTTCG	CAGCATCTTCATATACTACAGC	406	55
nleG2-3 (Z2149)	GGATGGAACCATACCTGG	CGCAATCAATTGCTAATGC	551	56
nleG5-2 (Z2151)	TGGAGGCTTTACGTCATGTCG	CCGGAACAAAGGGTTCACG	504	55
nleG6-2 (Z2150)	CGGGTCAGTGGATGATATGAGC	AAGTAGCATCTAGCGGTCGAGG	424	55
nleG9 (Z2560)	GTTCGTGCCCGAATTGTAGC	CACCAACCAAACGAGAAAATG	409	55
nleH1-2 (Z6021)	AACGCCTTATATTTTACC	AGCACAATTATCTCTTCC	589	52
nleH1-1 (Z0989)	GTTACCACCTTAAGTATCC	GTTTCTCATGAACACTCC	456	55
ent/*espL2*	GAATAACAATCACTCCTCACC	TTACAGTGCCCGATTACG	233	55

[a] Designations in parentheses are alternate gene names used previously.

bution of specific PAIs and *nle* genes to *E. coli* virulence is unknown in most cases.

Seropathotype classification was developed as an aid to assess the role of genomic islands in contributing to the public health risk associated with different STEC serotypes, especially those found in foods, animals, or the environment (15). Seropathotypes can provide the basis for a more precise molecular risk assessment (MRA) framework for STEC virulence in the public health setting. Genetic and epidemiological work has identified a number of genes in non-O157 STEC associated with virulence and severe disease in humans (38). These include the effectors *ent* (*espL2*/*Z4326*), *nleB*, and *nleE* (8, 32, 38) that are encoded on the O-Island 122 PAI and whose gene products are secreted by the LEE type III secretion system. The type III secretion system is a key genetic determinant of both colonization and persistence in nonhuman animal reservoirs and for virulence in humans after zoonotic transmission (6, 7, 10, 40). Genetic screens indicate that type III effectors contribute to persistence in cattle of both O157:H7 and non-O157 STEC strains (10, 34), and work in a mouse model of pathogenic *E. coli* links colonization, transmission, and virulence to type III effector genes common to attaching-and-effacing pathogens (37). Together, these data highlight a clinically relevant link between colonization in animals and human virulence that is influenced by the repertoire of type III secretion genes and assemblage of PAIs in the associated strain. This information has important implications for public health because the persistence of type III secretion-competent strains in animal populations increases the opportunities for zoonotic and environmental transmission of the most virulent strains. Identifying the type III secreted effectors that promote adaptation and persistence of STEC in animal reservoirs is therefore of major consequence.

The World Health Organization has called the rapid identification of virulent non-O157 STEC a public health priority (2). Since type III secretion facilitates the persistence, host-to-host spread, and virulence of STEC, our objective was to identify type III secreted virulence factors that distinguish non-O157 STEC strains posing a serious risk to humans from STEC strains that are not associated with severe and epidemic disease, a concept we call "molecular risk assessment" (i.e., MRA). We focused our efforts on the non-LEE effector genes encoded in genomic PAIs (O-Islands) (28, 29, 31) and analyzed the distribution of 16 *nle* genes among O157 and non-O157 STEC strains. We report the identification of 14 new *nle* genes in non-O157 STEC strains, grouped within three PAIs, that correlate independently with outbreak and HUS potential for humans. In addition, we show an *nle* gene dosing effect in non-O157 STEC, where strains associated with severe human disease have an increased number of *nle* genes. In a post-genomic era of emerging infectious diseases, MRA could be a strategic direction for public health initiatives responding to emerging bacteria in foods, animals, and the environment.

MATERIALS AND METHODS

Bacterial strains. 72 STEC isolates used in the present study are listed in Table S1 in the supplemental material. STEC strains of the same serotype are unique isolates from different patients or animals that were unlinked temporally and gave distinct macrorestriction enzyme digest patterns by pulsed-field gel electrophoresis. Seropathotype status was assigned to each STEC strain based on the reported frequencies of the serotype in outbreaks, HUS, and hemorrhagic colitis according to the methods of Karmali et al. (16). It should be noted that although strains EC2-032, EC3-480, EC96-371, and EC96-484 are of bovine origin, the serotypes they represent (O171:H2, O7:H4, O113:H4, and O172:NM, respectively) are associated with human diarrheal disease but not outbreaks or HUS.

Virulence gene content in non-O157 STEC strains. PCR was used to screen STEC strains for the presence of 16 genes whose gene products are substrates for the virulence-associated type III secretion system encoded in the LEE. The genes examined included the non-LEE-encoded effector genes (*nleA*, *nleB*, *nleB2*, *nleC*, *nleD*, *nleE*, *nleF*, *nleG*, *nleG2-1*, *nleG2-3*, *nleG5-2*, *nleG6-2*, *nleG9*, *nleH1*, *nleH2*, and *ent*/*espL2*) recently identified as type III secreted effectors by Deng et al. and Tobe et al. (8, 32). *nleB*, *nleE*, and *ent*/*espL2* have been reported elsewhere (38). The distribution in our strain collection of non-type III secretion-associated virulence genes (e.g., the verocytotoxin [VT] genes *VT1* and *VT2*, *hlyA*, *espP*, and *katP*) has been reported elsewhere (16). All PCR amplifications were carried out in 50-μl reaction mixtures containing PCR buffer (Perkin-Elmer Applied Biosystems, Foster City, CA), 250 μM deoxynucleoside triphosphates, 1 mM $MgCl_2$, 25 pmol of each primer, and 2 U of *Taq* DNA polymerase (AmpliTaq; Applied Biosystems). Cycling conditions for all *nle* genes consisted of an initial denaturation step at 94°C for 5 min, followed by 30 cycles of denaturation at 94°C for 30 s, annealing for 1 min, and elongation at 72°C for 2.5 min. Oligonucleotide sequences used for PCR are listed in Table 1. A given O-Island was defined as

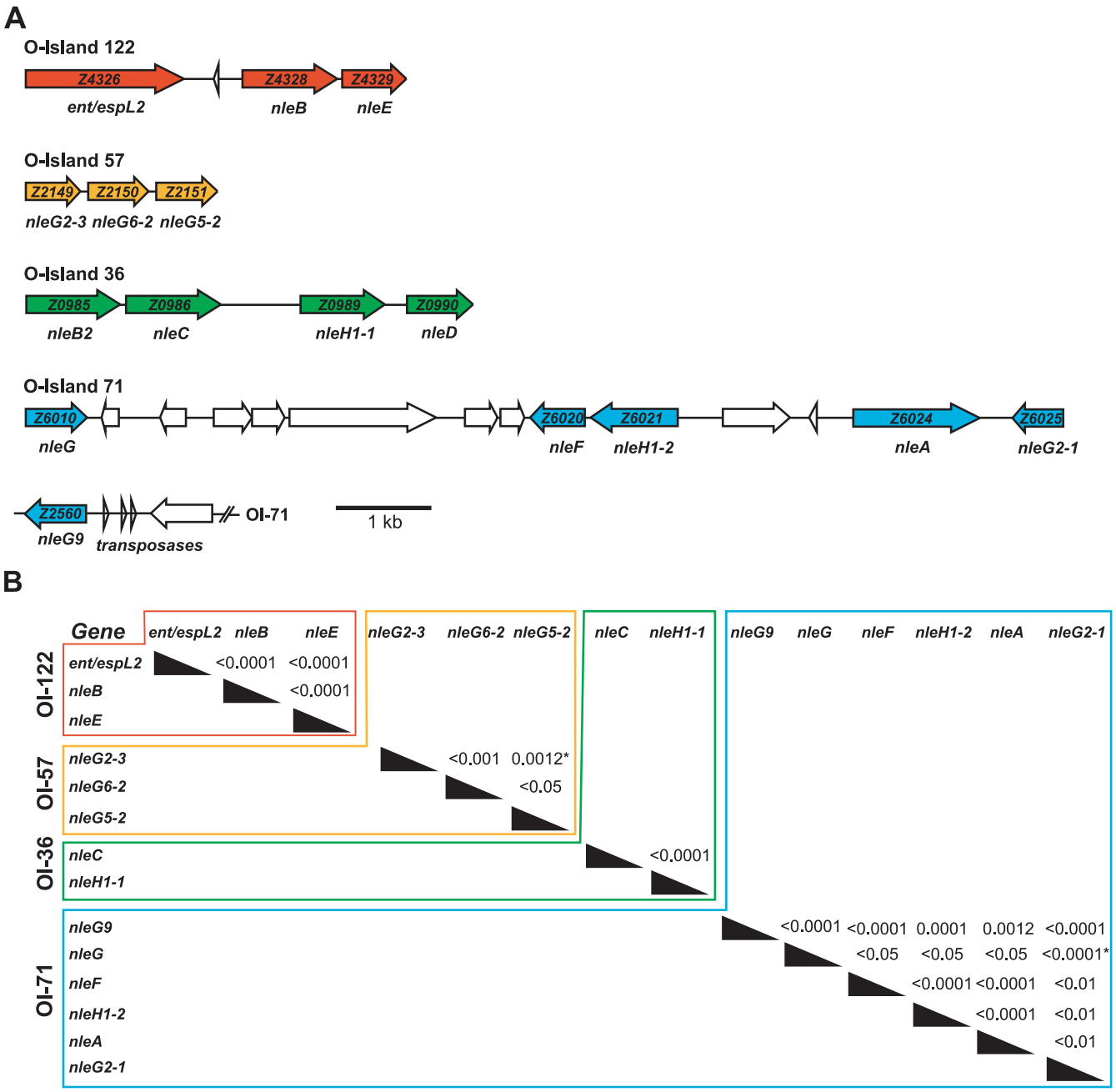

FIG. 1. Genetic organization of disease-associated type III effector *nle* genes in non-O157 STEC. (A) Graphic representation of the O-Island PAIs containing the *nle* effector genes examined in the present study in O157:H7 STEC strain EDL933. The direction of transcription for each open reading frame is indicated by an arrowhead. The type III effector genes examined are color-coded to each O-Island. *nleG9* is adjacent to OI-71 but separated by multiple transposase elements and was therefore not included as part of OI-71. The annotation of open reading frames within each O-Island was modified from Perna et al. (26). (B) Pairwise association of *nle* genes from O-Islands 122, 57, 36, and 71 in non-O157:H7 STEC strains linked to outbreaks. The data are *P* values (Fisher exact test) for the co-occurrence of each disease-associated *nle* gene within the linked O-Island. *nle* genes that did not associate with outbreaks or HUS by MRA (*nleB2* and *nleD*) were not included in pairwise comparisons. *, not-significant co-occurrence in non-HUS-associated strains (data not shown).

complete if all expected *nle* genes were present, whereas an O-Island lacking one or more *nle* genes was defined as incomplete.

Southern hybridizations. Strains that were initially PCR negative for an *nle* target were retested by Southern hybridization using previously published methods (16). Briefly, a DNA probe for each *nle* gene was amplified from O157:H7 using the primers listed in Table 1. Southern blots were performed using digoxigenin-labeled probes generated according to the manufacturer's instructions (Roche). Genomic DNA from test strains and control strains was extracted by using the DNeasy tissue kit from Qiagen (Hilden, Germany). Approximately 2 μg of DNA was digested with an excess of EcoRI and resolved on 0.6% agarose gels. The gel was then transferred to a nylon membrane (Roche Diagnostics, Mannheim, Germany), and hybridizations with labeled probes were carried out overnight at 42°C. Hybridized probes were detected by using digoxigenin substrate according to the manufacturer's instructions (Roche). EDL933 (26) and Sakai (41) O157:H7 strains were used as positive controls, and *E. coli* K-12 strain MG1655 (26) was used as a negative control. Strains that showed

TABLE 2. Distribution of complete O-Islands in HUS strains

Complete O-Island	Prevalence (%) within STEC group (no. of strains with O-Island/ total no. of strains)		Odds ratio (95% CI)[a]	*P*
	Non-HUS associated	HUS associated		
O-Island 36	7 (2/28)	14 (4/29)	2.1 (0.3–12.4)	NS[b]
O-Island 57	7 (2/28)	41 (12/29)	9.2 (1.8–46.3)	<0.01
O-Island 71	7 (2/28)	45 (13/29)	10.6 (2.1–53.1)	<0.01
O-Island 122	32 (9/28)	66 (19/29)	4.0 (1.3–12.1)	<0.05

[a] CI, confidence interval.
[b] NS, not significant.

TABLE 3. Distribution of complete O-Islands in outbreak strains

Complete O-Island	Prevalence (%) within STEC group (no. of strains with O-Island/ total no. of strains)		Odds ratio (95% CI)[a]	*P*
	Non-outbreak associated	Outbreak associated		
O-Island 36	7 (3/42)	20 (3/15)	3.3 (0.6–18.3)	NS[b]
O-Island 57	17 (7/42)	46 (7/15)	4.4 (1.2–16.0)	<0.05
O-Island 71	11 (5/42)	67 (10/15)	14.8 (3.6–61.4)	0.0001
O-Island 122	33 (14/42)	93 (14/15)	28 (3.3–235.2)	<0.0001

[a] CI, confidence interval.
[b] NS, not significant.

a positive signal by Southern blotting were defined as positive for that particular target.

Statistical analysis. The Fisher exact test for two-tailed significance was used to examine associations between HUS and outbreaks and the presence or absence of type III effector genes. MRA analysis (the ability of effector genes to predict HUS and outbreak-associated STEC) was performed by calculating the sensitivity, specificity, and predictive value for each *nle* gene examined. The prevalence of *nle* genes was compared in non-O157 STEC seropathotypes by using two-tailed unpaired *t* tests with Welch's correction when necessary.

RESULTS

Identification of *nle* genes associated with HUS and outbreaks from non-O157 STEC. We previously developed seropathotype categorization as a starting point to characterize highly virulent STEC-associated epidemic spread in the human population and severe disease, including HUS (16). However, the molecular basis for the increased virulence of some, but not all, non-O157 STEC strains is unknown. To address this, we took 72 clinical STEC isolates that collectively comprise all seropathotype categories (see Table S1 in the supplemental material) and examined in these strains the distribution of 16 *nle* genes carried in genomic O-Islands (28, 29, 31, 36) that encode putative or known virulence effectors secreted by the LEE type III secretion system (Fig. 1A). O-Islands are genetic regions absent from nonpathogenic *E. coli* and frequently contain virulence determinants. We first determined the prevalence of a complete O-Islands 36, 57, and 71 among strains associated with outbreaks or with HUS or not associated with severe disease. As mentioned, genetic content for O-Island 122 has been reported elsewhere (38). Complete O-Islands 57 and 71, but not O-Island 36, were more prevalent among strains causing HUS (Table 2) and outbreaks (Table 3) compared to strains not associated with epidemics or severe human disease, suggesting that O-Islands 57 and 71 may increase the virulence potential of STEC. Strains lacking complete PAIs were less frequently associated with human disease. The combined presence of incomplete O-Islands 57, 71, and 122 in a non-O157 STEC strain was significantly associated with an inability to cause outbreaks ($P < 0.0001$) and HUS ($P = 0.0173$). Interestingly, while strains possessing exactly one complete O-Island did not associate with human disease, strains with exactly two complete O-Islands were associated significantly with severe disease in humans ($P < 0.05$ for both HUS and outbreaks), and the presence of two or more complete O-Islands in a strain correlated with increased ability to cause severe disease in humans ($P < 0.001$ for both HUS and outbreaks), suggesting that PAIs may contribute additively to non-O157 STEC virulence.

Next, we examined individual *nle* effector genes for their prevalence in non-O157 STEC strains that were either associated or not associated with HUS and outbreaks. Among non-O157 STEC strains, the genes *nleA*, *nleB*, *nleC*, *nleE*, *nleF*, *nleG*, *nleG2-1*, *nleG2-3*, *nleG5-2*, *nleG6-2*, *nleG9*, *nleH1*, *nleH2*, and *ent*/*espL2* were each more prevalent in isolates associated with HUS after infection than in strains associated with only diarrhea or nonhuman animal infections (Table 4). The effectors *nleB2* and *nleD* did not show increased prevalence in HUS-associated non-O157 STEC strains and were similarly distributed between HUS-associated isolates and less-virulent strains. Similarly, STEC strains associated with outbreaks had a significantly higher prevalence of *nleA*, *nleB*, *nleC*, *nleE*, *nleF*, *nleG*, *nleG2-1*, *nleG2-3*, *nleG5-2*, *nleG6-2*, *nleG9*, *nleH1*, *nleH2*, and *ent*/*espL2* than did non-outbreak-associated strains (Table 5). The type III effector genes *nleB2* and *nleD* were equally distributed among outbreak isolates and less-virulent isolates of non-O157 STEC.

TABLE 4. Prevalence of *nle* genes in non-O157 STEC strains associated with HUS

Gene[a]	Prevalence (%) within STEC group (no. of strains with O-Island/ total no. of strains)		Odds ratio (95% CI)[b]	*P*
	Non-HUS associated	HUS associated		
nleA	32 (9/28)	69 (20/29)	4.7 (1.5–14.3)	0.0081
nleB	32 (9/28)	69 (20/29)	4.7 (1.5–14.3)	0.0081
nleB2	17 (5/28)	24 (7/29)	1.5 (0.4–5.3)	NS[c]
nleC	25 (7/28)	62 (18/29)	4.9 (1.6–15.3)	0.0074
nleD	7 (2/28)	14 (4/29)	2.1 (0.3–12.4)	NS
nleE	32 (9/28)	66 (19/29)	4.0 (1.3–12.1)	0.0173
nleF	32 (9/28)	62 (18/29)	3.5 (1.2–10.3)	0.0343
nleG (Z6010)	11 (3/28)	62 (18/29)	13.6 (3.3–56.1)	<0.0001
nleG2-1 (Z6025)	18 (5/28)	62 (18/29)	7.5 (2.2–25.6)	0.0011
nleG2-3 (Z2149)	25 (7/28)	66 (19/29)	5.7 (1.8–18)	0.0033
nleG5-2 (Z2151)	7 (2/28)	41 (12/29)	9.1 (1.8–46)	0.0046
nleG6-2 (Z2150)	18 (5/28)	52 (15/29)	4.9 (1.5–16.5)	0.0119
nleG9 (Z2560)	7 (2/28)	45 (13/29)	10.6 (2.1–53)	0.002
nleH1-2 (Z6021)	32 (9/28)	62 (18/29)	3.5 (1.2–10.3)	0.0343
nleH1-1 (Z0989)	32 (9/28)	62 (18/29)	3.5 (1/2–10.3)	0.0343
ent/*espL2*	32 (9/28)	69 (20/29)	4.7 (1.5–14.3)	0.0081

[a] Designations in parentheses are alternate gene names used previously.
[b] CI, confidence interval.
[c] NS, not significant.

TABLE 5. Prevalence of *nle* genes in non-O157 STEC strains associated with outbreaks

Gene[a]	Prevalence (%) within STEC group (no. of strains with O-Island/ total no. of strains)		Odds ratio (95% CI)[b]	*P*
	Non-outbreak associated	Outbreak-associated		
nleA	33 (14/42)	100 (15/15)	61 (3.4–1,093)	<0.0001
nleB	33 (14/42)	100 (15/15)	61 (3.4–1,093)	<0.0001
nleB2	14 (6/42)	40 (6/15)	4 (1–15.4)	NS[c]
nleC	29 (12/42)	87 (13/15)	16 (3.2–83)	0.0001
nleD	7 (3/42)	20 (3/15)	3.3 (0.6–18.3)	NS
nleE	33 (14/42)	93 (14/15)	28 (3.3–235.2)	<0.0001
nleF	33 (14/42)	87 (13/15)	13 (2.6–65.8)	0.0006
nleG (Z6010)	19 (8/42)	87 (13/15)	28 (5.2–147.7)	<0.0001
nleG2-1 (Z6025)	24 (10/42)	87 (13/15)	21 (4–108.3)	<0.0001
nleG2-3 (Z2149)	29 (12/42)	93 (14/15)	35 (4.1–296.6)	<0.0001
nleG5-2 (Z2151)	17 (7/42)	46 (7/15)	4.4 (1.2–16)	0.0344
nleG6-2 (Z2150)	24 (10/42)	67 (10/15)	6.4 (1.8–23.2)	0.0046
nleG9 (Z2560)	11 (5/42)	67 (10/15)	14.8 (3.6–61.4)	0.0001
nleH1-2 (Z6021)	33 (14/42)	87 (13/15)	13 (2.6–65.8)	0.0006
nleH1-1 (Z0989)	31 (13/42)	93 (14/15)	31 (3.7–263.4)	<0.0001
ent/espL2	33 (14/42)	100 (15/15)	60.9 (3.4–1,093)	<0.0001

[a] Designations in parentheses are alternate gene names used previously.
[b] CI, confidence interval.
[c] NS, not significant.

Distribution of type III effectors in strains from each seropathotype class. The association of the non-LEE effectors with each other was examined for each O-Island across outbreak-associated non-O157 STEC serotypes (Fig. 1B). In each case, type III effectors within the same O-Island co-associate with each other, suggesting that the genes within O-Islands 122, 57, 36, and 71 are likely transferred en bloc by horizontal gene transfer. The question of whether or not virulence genes contribute additively to disease caused by STEC is an important question suggested by work in other pathogenic bacteria (reviewed in reference 13). To begin to address this, we determined whether the cumulative number of disease-associated *nle* genes within a non-O157 STEC strain correlated with the propensity of that strain to cause severe disease in humans. The distribution of *nle* gene content across the study strains ($n = 72$) is presented in Fig. 2A. All O157:H7 strains examined (representing seropathotype A) contained all of the disease-associated *nle* genes identified (14 genes out of a total of 16 examined), a finding consistent with previous work examining O157:H7 gene content (36). Although O157:H7 was not our primary focus for this work, it is an important benchmark against which comparisons to non-O157 strains can be made. In contrast, non-O157 STEC strains contain a variable repertoire of *nle* virulence genes that correlates with their clinical virulence potential. Figure 2A shows the cumulative distribution of *nle* gene content across the seropathotype categories. A greater proportion of seropathotype B (73.3%) and seropathotype C (35%) strains contained more than 13 disease-associated *nle* genes compared to seropathotype D (14%) and seropathotype E (0%) strains. We then examined the mean *nle* effector gene content of all of the non-O157 STEC strains linked to epidemics and HUS (Fig. 2B). Of the effector genes examined, STEC strains linked to outbreaks contained significantly more effector genes than STEC strains not associated with outbreaks (outbreak strains, 11.7 ± 0.7 genes; nonoutbreak strains, 3.8 ± 0.9 genes; $P = 0.0001$ [unpaired Student *t* test with Welch's correction]). Strains linked to HUS also contained a greater number of *nle* virulence genes than did non-HUS strains (Fig. 2B) (HUS strains, 8.4 ± 1.1 genes; non-HUS strains, 3.3 ± 0.9 genes; $P = 0.0014$ [unpaired *t* test]). These data suggest that virulence factors, such as type III effectors, may contribute additively to non-O157 STEC virulence. An analysis of the relative contribution of each genetic element to

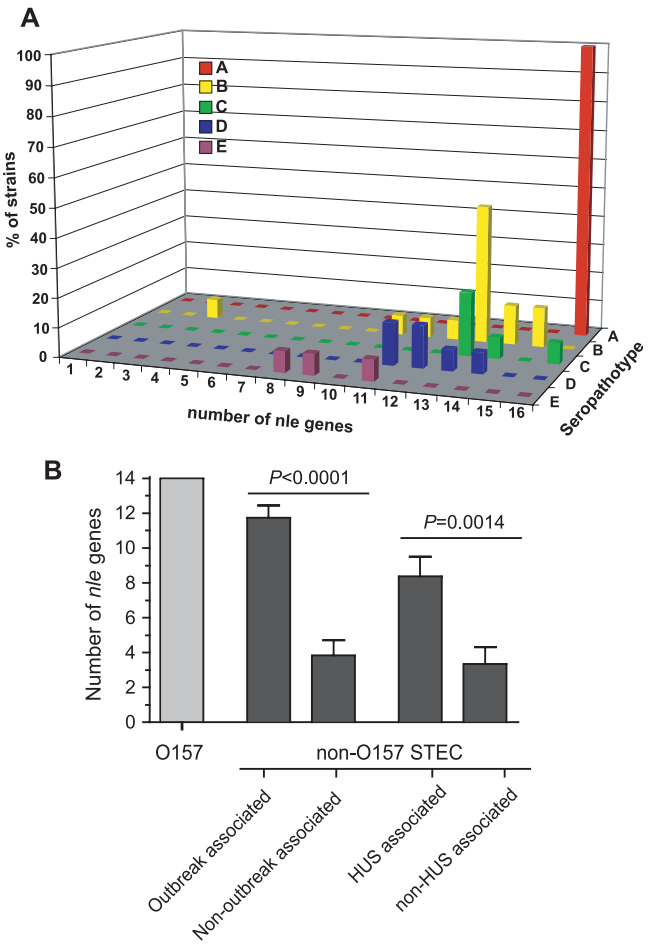

FIG. 2. Type III effector-gene dosage correlates with STEC virulence. (A) Distribution of *nle* gene content in O157 strains ($n = 15$) and non-O157 STEC strains ($n = 57$) (see Table S1 in the supplemental material for strain details). All O157 strains examined (red) contain all of the *nle* genes presented in Table 1. Non-O157 strains linked to human disease (yellow, green, and blue) contain a variable repertoire of disease-associated *nle* genes. Seropathotype E STEC strains not associated with human disease (purple) contain fewer *nle* genes relative to strains that are associated with outbreaks and HUS (seropathotype B), HUS only (seropathotype C), and diarrhea (seropathotype D). The data shown are the percentage of strains positive for the number of *nle* genes within each seropathotype. (B) Prevalence of *nle* genes in O157 and non-O157 STEC strains. Epidemic non-O157 strains have a significantly higher *nle* gene content compared to strains not linked to epidemics. Similarly, HUS-associated non-O157 strains have a significantly higher *nle* gene content compared to non-HUS-associated strains. O157 STEC contains all disease-associated *nle* genes identified. The data are means with standard errors.

virulence can therefore be addressed with a larger sample of STEC strains (23).

DISCUSSION

The emergence of non-O157 STEC in severe and epidemic human disease is of great global concern. In the postgenomic era, comparative genomics coupled with an evaluation of genomic island content and an assessment of the virulence genes these islands encode has opened up new tools to assess the public health risks associated with non-O157 STEC strains from food, animals, and the environment. We call this concept MRA. Although the toxins VT1 and VT2 are major STEC virulence factors, the presence of these toxins was equally distributed among the different seropathotypes, including those not associated with human disease. Therefore, these markers seem unsuitable for differentiating strains with high virulence potential to humans. As a first step toward STEC-MRA realization, we have identified 14 genes in non-O157 STEC that are more prevalent in strains associated with epidemic spread and hemolytic-uremic syndrome in the human population. Furthermore, we present data that are consistent with additive *nle* gene content contributing to increased burden of disease. Our collection of STEC seropathotypes is currently 72 strains and thus a major focus of future MRA efforts will be to increase the non-O157 STEC strain collection size to further refine the relative contributions to virulence of the genes identified in this work as high-risk (23). In the future, such efforts could be coordinated through national and international surveillance programs.

The genes identified here encode substrates of a type III secretion system in pathogenic *E. coli*. A great deal of data links the type III secretion system to STEC colonization and persistence in ruminant reservoirs, a major nonpathogenic niche for STEC. The association of *nleA*, *nleB*, *nleE*, and *ent*/*espL2* with severe disease in humans is consistent with their demonstrable roles in virulence in the closely related attaching-and-effacing mouse pathogen *Citrobacter rodentium*, which causes severe morbidity and mortality in mice (11, 18, 24, 38, 39). For example, deletion of *nleA* (*espI*) from *C. rodentium* prevents the lethal infection of mice (11) and reduces the virulence of the strain in competitive infections with wild-type *C. rodentium* by ~5 orders of magnitude (24). Similarly, in the *C. rodentium* mouse model, *nleB* is required for full colonization of the mouse colon and contributes to colonic hyperplasia (18, 38). Mutants lacking *nleB* are unable to cause mortality in C3H/HeJ mice (38), indicating that *nleB* encodes an important virulence determinant that plays a central role in the disease of attaching-and-effacing pathogens. The *nle* genes *nleH1-2* and *nleF* were also identified here as being associated with epidemic potential and severe human disease. Of relevance, whole-genome analyses have distinguished two lineages of O157:H7 *E. coli* into which human-specific (lineage I) and bovine-specific (lineage II) isolates are nonrandomly distributed (19). A recent comparative genomic hybridization study of 31 O157:H7 *E. coli* strains (42) showed that *nleH1-2* and *nleF* were associated with lineage I strains of O157:H7 that are linked to human disease and were absent from lineage II strains. Together, these data suggest that *nleH1-2* and *nleF* may increase the virulence potential of non-O157 STEC in humans or that strains bearing these genes are more readily transmitted to humans from bovine sources. That *nleB2* and *nleD* were not more prevalent in STEC strains linked to HUS and/or epidemics suggests that their contribution to human disease may be more subtle or perhaps functionally overlapping, a notion that is supported by virulence data for calves (22). Although *nleC* had no quantifiable role in virulence in lambs or calves, it appears to potentiate pathogenesis during *C. rodentium* infection of mice (M. Wickham and B. Finlay, unpublished data). The question remains whether *nleC* contributes to STEC colonization or persistence in cattle, and further work in determining the role of *nleC* in maintaining animal reservoirs is warranted given its association with severe disease in humans. Similarly, further study is required to determine the precise role of the disease-associated *nle* genes in non-O157 STEC virulence and ecology.

The connection between STEC persistence in ruminants and virulence in humans is noteworthy. Selection pressure in a nonpathogenic (commensal) setting that selects for type III-competent STEC variants would contemporaneously increase the likelihood of zoonotic transmission of more-virulent strains to humans. This notion is supported by recent data from EnterNet, a global surveillance consortium of 35 countries that track enteric infectious diseases. Cases of human disease caused by non-O157 STEC increased globally by 60.5% between 2000 and 2005, while at the same time cases caused by O157 STEC increased by only 13% (1). Among the top five non-O157 serotypes most frequently causing human disease in 2005, 80% belong to seropathotype B and 20% belong to seropathotype C (1), and none belong to the less-virulent seropathotypes D and E, suggesting that selection for more-virulent strains is currently taking place.

In regard to transmission dynamics, one of the characteristic features of O157:H7 STEC is a very low infectious dose, estimated to be 100 to 200 bacteria (25). We recently reported that NleB decreases the infectious dose required for colonization and disease in the *C. rodentium* mouse model of attaching-and-effacing pathogens (38). Given that *nleB* was present in all seropathotype A and B strains examined and hence linked to HUS and epidemic spread, it is possible that *nleB* modulates the infectivity of STEC in the human population as well. In addition, attaching-and-effacing enteropathogens lacking some *nle* genes are unable to cause severe morbidity or mortality in mice (11, 24, 39), suggesting that a concerted action of type III effectors contributes to non-O157 STEC pathogenicity. Using the attaching-and-effacing mouse pathogen, *C. rodentium*, we showed that type III secreted effectors had quantifiable and distinctive roles in host-to-host transmission success (37), providing methodology to examine other putative virulence determinants in the attaching-and-effacing pathogen complex. In one disease-associated non-O157 STEC strain, the presence of few *nle* genes suggests that additional virulence genes, in addition to those examined here, are important to the pathogenic potential of non-O157 STEC. Further genomics work to identify and characterize such virulence determinants is required. We believe that new applications in public health pathogenomics will depend on identifying high-risk strains based on a better understanding of what determines virulence. In this way, instructive genomic information can be built into analytical approaches for surveillance and epidemiology even before a

full understanding of the molecular mechanisms of pathogenesis is realized (27, 35). For example, it is now feasible to develop pathogen diagnostic tests based on multiplex nucleic acid amplification and microfluidics-based detection on standardized platforms for use in hospital service labs or public health laboratories (21). In addition, DNA arrays can be used to examine the gene inventory from clinical strains of various bacteria including STEC (42), offering a genetic barcoding strategy. Such approaches present feasible platforms for MRA implementation using customized probes or chips for virulence genes and whole PAIs linked to human disease. Future work examining the role in STEC virulence, transmission, and persistence of the disease-associated genes described here will facilitate a better understanding of the contribution of these genes to the population burden of disease. An MRA strategy that serves to identify genes or whole PAIs linked to human disease has tremendous public health implications for identifying high-risk STEC at the interfaces of humans, animals, and the environment.

ACKNOWLEDGMENTS

We thank all members of the Coombes laboratory for critical assessment of the manuscript.

B.K.C. is the recipient of a New Investigator Award from the Canadian Institutes of Health Research (CIHR) and a Young Investigator Award from the American Society for Microbiology. M.E.W. was a CJ Martin Fellow of the National Health and Medical Research Council of Australia and was supported by postdoctoral fellowships from the CIHR, the Michael Smith Foundation for Health Research, and The Killam Trusts. B.B.F. is a Howard Hughes International Research Scholar and the UBC Peter Wall Distinguished Professor. This study was funded by an operating grant from the CIHR Institute of Infection and Immunity Safe Food and Water Initiative and operating grants from the CIHR (B.K.C. and B.B.F.), the HHMI (B.B.F.), and the Public Health Agency of Canada (B.K.C.).

REFERENCES

1. **Anonymous.** 2005. European Commission Annual Report 2005: surveillance of enteric pathogens in Europe and beyond; 1786/2002/EC. International surveillance network for the enteric infections–*Salmonella*, VTEC O157 and *Campylobacter*. European Commission, Brussels, Belgium.
2. **Anonymous.** 1998. Zoonotic non-O157 Shiga toxin-producing *Escherichia coli* (STEC). Report of a WHO Scientific Workshop Group Meeting WHO/CSR/APH/98.8. World Health Organization, Geneva, Switzerland.
3. **Anonymous.** 2000. Waterborne outbreak of gastroenteritis associated with a contaminated municipal water supply, Walkerton, Ontario, May-June 2000. Can. Commun. Dis. Rep. **26:**170–173.
4. **Brooks, J. T., E. G. Sowers, J. G. Wells, K. D. Greene, P. M. Griffin, R. M. Hoekstra, and N. A. Strockbine.** 2005. Non-O157 Shiga toxin-producing *Escherichia coli* infections in the United States, 1983–2002. J. Infect. Dis. **192:**1422–1429.
5. **Caprioli, A., A. E. Tozzi, G. Rizzoni, and H. Karch.** 1997. Non-O157 Shiga toxin-producing *Escherichia coli* infections in Europe. Emerg. Infect. Dis. **3:**578–579.
6. **Cornick, N. A., S. L. Booher, and H. W. Moon.** 2002. Intimin facilitates colonization by *Escherichia coli* O157:H7 in adult ruminants. Infect. Immun. **70:**2704–2707.
7. **Dean-Nystrom, E. A., B. T. Bosworth, W. C. Cray, Jr., and H. W. Moon.** 1997. Pathogenicity of *Escherichia coli* O157:H7 in the intestines of neonatal calves. Infect. Immun. **65:**1842–1848.
8. **Deng, W., J. L. Puente, S. Gruenheid, Y. Li, B. A. Vallance, A. Vazquez, J. Barba, J. A. Ibarra, P. O'Donnell, P. Metalnikov, K. Ashman, S. Lee, D. Goode, T. Pawson, and B. B. Finlay.** 2004. Dissecting virulence: systematic and functional analyses of a pathogenicity island. Proc. Natl. Acad. Sci. USA **101:**3597–3602.
9. **Dobrindt, U., B. Hochhut, U. Hentschel, and J. Hacker.** 2004. Genomic islands in pathogenic and environmental microorganisms. Nat. Rev. Microbiol. **2:**414–424.
10. **Dziva, F., P. M. van Diemen, M. P. Stevens, A. J. Smith, and T. S. Wallis.** 2004. Identification of *Escherichia coli* O157: H7 genes influencing colonization of the bovine gastrointestinal tract using signature-tagged mutagenesis. Microbiology **150:**3631–3645.
11. **Gruenheid, S., I. Sekirov, N. A. Thomas, W. Deng, P. O'Donnell, D. Goode, Y. Li, E. A. Frey, N. F. Brown, P. Metalnikov, T. Pawson, K. Ashman, and B. B. Finlay.** 2004. Identification and characterization of NleA, a non-LEE-encoded type III translocated virulence factor of enterohaemorrhagic *Escherichia coli* O157:H7. Mol. Microbiol. **51:**1233–1249.
12. **Hacker, J., and J. B. Kaper.** 2000. Pathogenicity islands and the evolution of microbes. Annu. Rev. Microbiol. **54:**641–679.
13. **Johnson, J. R.** 1991. Virulence factors in *Escherichia coli* urinary tract infection. Clin. Microbiol. Rev. **4:**80–128.
14. **Johnson, K. E., C. M. Thorpe, and C. L. Sears.** 2006. The emerging clinical importance of non-O157 Shiga toxin-producing *Escherichia coli*. Clin. Infect. Dis. **43:**1587–1595.
15. **Karmali, M. A.** 2004. Infection by Shiga toxin-producing *Escherichia coli*: an overview. Mol. Biotechnol. **26:**117–122.
16. **Karmali, M. A., M. Mascarenhas, S. Shen, K. Ziebell, S. Johnson, R. Reid-Smith, J. Isaac-Renton, C. Clark, K. Rahn, and J. B. Kaper.** 2003. Association of genomic O island 122 of *Escherichia coli* EDL 933 with verocytotoxin-producing *Escherichia coli* seropathotypes that are linked to epidemic and/or serious disease. J. Clin. Microbiol. **41:**4930–4940.
17. **Karmali, M. A., M. Petric, C. Lim, P. C. Fleming, G. S. Arbus, and H. Lior.** 1985. The association between idiopathic hemolytic-uremic syndrome and infection by verotoxin-producing *Escherichia coli*. J. Infect. Dis. **151:**775–782.
18. **Kelly, M., E. Hart, R. Mundy, O. Marches, S. Wiles, L. Badea, S. Luck, M. Tauschek, G. Frankel, R. M. Robins-Browne, and E. L. Hartland.** 2006. Essential role of the type III secretion system effector NleB in colonization of mice by *Citrobacter rodentium*. Infect. Immun. **74:**2328–2337.
19. **Kim, J., J. Nietfeldt, and A. K. Benson.** 1999. Octamer-based genome scanning distinguishes a unique subpopulation of *Escherichia coli* O157:H7 strains in cattle. Proc. Natl. Acad. Sci. USA **96:**13288–13293.
20. **Lawrence, J. G.** 2005. Common themes in the genome strategies of pathogens. Curr. Opin. Genet. Dev. **15:**584–588.
21. **Mahony, J., S. Chong, F. Merante, S. Yaghoubian, T. Sinha, C. Lisle, and R. Janeczko.** 2007. Development of a respiratory virus panel test for detection of twenty human respiratory viruses by use of multiplex PCR and a fluid microbead-based assay. J. Clin. Microbiol. **45:**2965–2970.
22. **Marches, O., S. Wiles, F. Dziva, R. M. La Ragione, S. Schuller, A. Best, A. D. Phillips, E. L. Hartland, M. J. Woodward, M. P. Stevens, and G. Frankel.** 2005. Characterization of two non-locus of enterocyte effacement-encoded type III-translocated effectors, NleC and NleD, in attaching and effacing pathogens. Infect. Immun. **73:**8411–8417.
23. **McClelland, E. E., P. Bernhardt, and A. Casadevall.** 2006. Estimating the relative contributions of virulence factors for pathogenic microbes. Infect. Immun. **74:**1500–1504.
24. **Mundy, R., L. Petrovska, K. Smollett, N. Simpson, R. K. Wilson, J. Yu, X. Tu, I. Rosenshine, S. Clare, G. Dougan, and G. Frankel.** 2004. Identification of a novel *Citrobacter rodentium* type III secreted protein, EspI, and roles of this and other secreted proteins in infection. Infect. Immun. **72:**2288–2302.
25. **Nataro, J. P., and J. B. Kaper.** 1998. Diarrheagenic *Escherichia coli*. Clin. Microbiol. Rev. **11:**142–201.
26. **Perna, N. T., G. Plunkett III, V. Burland, B. Mau, J. D. Glasner, D. J. Rose, G. F. Mayhew, P. S. Evans, J. Gregor, H. A. Kirkpatrick, G. Posfai, J. Hackett, S. Klink, A. Boutin, Y. Shao, L. Miller, E. J. Grotbeck, N. W. Davis, A. Lim, E. T. Dimalanta, K. D. Potamousis, J. Apodaca, T. S. Anantharaman, J. Lin, G. Yen, D. C. Schwartz, R. A. Welch, and F. R. Blattner.** 2001. Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7. Nature **409:**529–533.
27. **Pompe, S., J. Simon, P. M. Wiedemann, and C. Tannert.** 2005. Future trends and challenges in pathogenomics: a foresight study. EMBO Rep. **6:**600–605.
28. **Shen, S., M. Mascarenhas, R. Morgan, K. Rahn, and M. A. Karmali.** 2005. Identification of four fimbria-encoding genomic islands that are highly specific for verocytotoxin-producing *Escherichia coli* serotype O157 strains. J. Clin. Microbiol. **43:**3840–3850.
29. **Shen, S., M. Mascarenhas, K. Rahn, J. B. Kaper, and M. A. Karmali.** 2004. Evidence for a hybrid genomic island in verocytotoxin-producing *Escherichia coli* CL3 (serotype O113:H21) containing segments of EDL933 (serotype O157:H7) O islands 122 and 48. Infect. Immun. **72:**1496–1503.
30. **Tarr, P. I., and M. A. Neill.** 1996. Perspective: the problem of non-O157:H7 Shiga toxin (verocytotoxin)-producing *Escherichia coli*. J. Infect. Dis. **174:**1136–1139.
31. **Taylor, D. E., M. Rooker, M. Keelan, L. K. Ng, I. Martin, N. T. Perna, N. T. Burland, and F. R. Blattner.** 2002. Genomic variability of O islands encoding tellurite resistance in enterohemorrhagic *Escherichia coli* O157:H7 isolates. J. Bacteriol. **184:**4690–4698.
32. **Tobe, T., S. A. Beatson, H. Taniguchi, H. Abe, C. M. Bailey, A. Fivian, R. Younis, S. Matthews, O. Marches, G. Frankel, T. Hayashi, and M. J. Pallen.** 2006. An extensive repertoire of type III secretion effectors in *Escherichia coli* O157 and the role of lambdoid phages in their dissemination. Proc. Natl. Acad. Sci. USA **103:**14941–14946.
33. **Tozzi, A. E., A. Caprioli, F. Minelli, A. Gianviti, L. De Petris, A. Edefonti, G. Montini, A. Ferretti, T. De Palo, M. Gaido, and G. Rizzoni.** 2003. Shiga toxin-producing *Escherichia coli* infections associated with hemolytic-uremic syndrome, Italy, 1988–2000. Emerg. Infect. Dis. **9:**106–108.

34. **van Diemen, P. M., F. Dziva, M. P. Stevens, and T. S. Wallis.** 2005. Identification of enterohemorrhagic *Escherichia coli* O26:H− genes required for intestinal colonization in calves. Infect. Immun. **73:**1735–1743.
35. **Weinstock, G. M.** 2000. Genomics and bacterial pathogenesis. Emerg. Infect. Dis. **6:**496–504.
36. **Wick, L. M., W. Qi, D. W. Lacher, and T. S. Whittam.** 2005. Evolution of genomic content in the stepwise emergence of *Escherichia coli* O157:H7. J. Bacteriol. **187:**1783–1791.
37. **Wickham, M. E., N. F. Brown, E. C. Boyle, B. K. Coombes, and B. B. Finlay.** 2007. Virulence is positively selected by transmission success between mammalian hosts. Curr. Biol. **17:**783–788.
38. **Wickham, M. E., C. Lupp, M. Mascarenhas, A. Vazquez, B. K. Coombes, N. F. Brown, B. A. Coburn, W. Deng, J. L. Puente, M. A. Karmali, and B. B. Finlay.** 2006. Bacterial genetic determinants of non-O157 STEC outbreaks and hemolytic-uremic syndrome after infection. J. Infect. Dis. **194:**819–827.
39. **Wickham, M. E., C. Lupp, A. Vazquez, M. Mascarenhas, B. Coburn, B. K. Coombes, M. A. Karmali, J. L. Puente, W. Deng, and B. B. Finlay.** 2007. *Citrobacter rodentium* virulence in mice associates with bacterial load and the type III effector NleE. Microbes Infect. **9:**400–407.
40. **Woodward, M. J., A. Best, K. A. Sprigings, G. R. Pearson, A. M. Skuse, A. Wales, C. M. Hayes, J. M. Roe, J. C. Low, and R. M. La Ragione.** 2003. Non-toxigenic *Escherichia coli* O157:H7 strain NCTC12900 causes attaching-effacing lesions and *eae*-dependent persistence in weaned sheep. Int. J. Med. Microbiol. **293:**299–308.
41. **Yokoyama, K., K. Makino, Y. Kubota, M. Watanabe, S. Kimura, C. H. Yutsudo, K. Kurokawa, K. Ishii, M. Hattori, I. Tatsuno, H. Abe, M. Yoh, T. Iida, M. Ohnishi, T. Hayashi, T. Yasunaga, T. Honda, C. Sasakawa, and H. Shinagawa.** 2000. Complete nucleotide sequence of the prophage VT1-Sakai carrying the Shiga toxin 1 genes of the enterohemorrhagic *Escherichia coli* O157:H7 strain derived from the Sakai outbreak. Gene **258:**127–139.
42. **Zhang, Y., C. Laing, M. Steele, K. Ziebell, R. Johnson, A. K. Benson, E. Taboada, and V. P. Gannon.** 2007. Genome evolution in major *Escherichia coli* O157:H7 lineages. BMC Genomics **8:**121.