(CS1 + CS3 (± CS21)) and L2 (CS2 + CS3 (± CS21)) were positive for LT + STh and shared the same LT and STh allelic variants. Interestingly, this LT allele was not found in any other isolates of this data set, whereas the ST variant (STa3/4) was found elsewhere in the tree (L5). Two other lineages, L3 (CFA/I) and L5 (CS5 + CS6), comprised LT + STh and STh-only isolates. Isolates in both L3 and L5 had the same translated LT variant, LT2 (ref. 17). These results demonstrate that not only the colonization factor profile but also the toxin allele profile is associated with chromosomal background, further supporting the idea that specific chromosomal-plasmid combinations are present in stable ETEC lineages.

Specific plasmid incompatibility groups are found in each lineage

Using *in silico* PCR with specific primers for the most common plasmid incompatibility groups, we could estimate the number of different replicons present in each ETEC isolate and relate this number to the virulence factors present in each of the lineages. Sequences corresponding to plasmid incompatibility groups could be identified in 317 plasmids from the 362 ETEC isolates sequenced. Isolates in L1 (CS1 + CS3) contained 2–5 different incompatibility groups, FII, FIIY, FrepB, FIB and I1 (**Supplementary Figs. 1** and **2**, and **Supplementary Table 2**), corresponding to the pCoo plasmid encoding CS1 and incompatibility groups I1 (CS1) and FII (CS3) identified in the ETEC E24377A (CS1 + CS3) strain[19,20]. FIB, known to be associated with virulence traits[20], was identified in all isolates within the L3 (CFA/I) lineage. Additional incompatibility groups were found in the majority of the isolates in lineage L3; among these were FII, FIIY and FrepB (**Supplementary Figs. 1** and **2**, and **Supplementary Table 2**). In the fully annotated genome of the reference strain H10407 (CFA/I), four plasmids have previously been identified: two larger plasmids with FII, one of which harbors the CFA/I operon, and two smaller plasmids with ColE1 (ref. 21). Although the number of incompatibility groups varied within a lineage, a pattern of conserved incompatibility groups in each lineage was apparent.

Five selected ETEC lineages emerged during the last century

Our findings demonstrate a clear association between the O antigen and virulence profiles, which correlates well with the chromosomal genotype of the ETEC lineages. To estimate the time of emergence of the selected lineages, we reconstructed an evolutionary history of these lineages using BEAST. A maximum-clade-credibility (MCC) tree was generated for each lineage, and this tree was used to estimate the most recent common ancestor (MRCA). The substitution rates (number of substitutions per site per year) for the five selected lineages were largely consistent and were estimated to be 1.0×10^{-6} (L1), 1.0×10^{-6} (L2), 3.7×10^{-7} (L3), 4.0×10^{-7} (L4) and 1.1×10^{-6} (L5). The substitution rates in L1, L2 and L5 were similar to the mutation rate of 1.57×10^{-6} estimated in *Streptococcus pneumoniae*[22]. Lineages L3 and L4 had slightly lower substitution rates, which were in the same range as that of *Clostridium difficile*[23]. Estimates for L1–L5 correspond to an accumulation of approximately 2–5.5 SNPs per genome per year.

In a time-dependent reconstruction of the identified lineages, L1–L5 were estimated to have emerged between 51 and 174 years ago (**Fig. 2**). Hence, these lineages emerged between the 1840s and 1970s. This result further supports the idea that the lineages are stable and implies a tighter and longer-term coupling of the chromosome and plasmid than has previously been appreciated.

DISCUSSION

ETEC infections are a major cause of diarrhea in low- and middle-income countries, and, although diarrhea incidence may be declining slightly[24], the total burden in children in these countries remains high. Additional research is needed to improve prevention as well as treatment; hence, a more detailed understanding of what actually constitutes a naturally occurring ETEC strain is vital. We have applied high-resolution genomic analysis to determine the phylogenetic structure of a globally representative ETEC collection in the context of the whole *E. coli* species. ETEC isolates were found across *E. coli* phylogroups A, B1, B2, D and E in the MCG-based phylogenetic tree, in agreement with previous multilocus sequence type (MLST)-based studies (using internal sequences of housekeeping genes: *adk*, *fumC*, *gyrB*, *icd*, *mdh*, *purA* and *recA*)[8,25], demonstrating that ETEC are indeed genetically diverse (**Fig. 1**). Despite this diversity, we unexpectedly found that many ETEC isolates formed discrete lineages. Combining the colonization factor and toxin profiles of the ETEC isolates with whole-genome data, we demonstrated the existence of clearly identifiable ETEC lineages. These lineages were not only phylogenetically related but also shared consistent plasmid-encoded virulence profiles. For example, the closely related L1 and L2 lineages harbored CS1 + CS3 (± CS21) and CS2 + CS3 (± CS21) profiles, respectively, and shared specific LT and ST alleles, and they all expressed the O6 antigen (**Fig. 2**, **Table 1** and **Supplementary Figs. 1** and **2**). This close clustering of isolates from different parts of the world collected over a period of 30 years may suggest that the acquisition of plasmid-encoded virulence factors occurred once and was then followed by a clonal expansion of isolates carrying the same virulence profile. It is clear that this event has not only occurred in a single case: we observed multiple distantly related lineages where the same pattern was evident. For instance, lineage L5 mainly comprised isolates positive for CS5 + CS6 that were not found elsewhere in the data set. In contrast, CFA/I-positive isolates were found in two lineages that had different O antigens but shared the same LT allele. Isolates expressing CS6 alone or in combination with CS8 or CS21 were identified in four lineages (L4 and L7–L9) spread across the MCG-based phylogenetic tree as well as in additional lineages together with colonization factor–negative isolates (**Figs. 1** and **2**, and **Supplementary Figs. 1** and **2**). CS6 is known to be a diverse colonization factor, as several variants of the genes encoding the structural subunits, *cssA* and *cssB*, have been identified[26]. CS6-positive isolates clustered on the basis of different variants of the CS6 structural subunits (data not shown). Screening for plasmid incompatibility groups provides further evidence that a specific plasmid and chromosomal background is stably maintained in a population (**Supplementary Figs. 1** and **2**). In summary, ETEC isolates with distinct virulence profiles can form monophyletic lineages, for example, CS1 + CS3 (± CS21), CS2 + CS3 (± CS21) and CS5 + CS6 isolates, or arise in isolates with different chromosomal backgrounds, for example, CFA/I and CS6 isolates.

It has previously been suggested that the acquisition of colonization factor and toxin genes is sufficient for the emergence of pathogenic ETEC isolates[8]. Previous studies had indicated a low frequency of phylogenetic clustering of ETEC isolates[9,10] and that the acquisition of virulence-related genes occurred at multiple times, and no common clonal lineages with distinct virulence profiles had been identified[8,12,27]. Indeed, a portion of the isolates in our ETEC collection form lineages with isolates of mixed O antigens and virulence profiles; however, these isolates are mostly represented by colonization factor–negative isolates and by ETEC expressing less prevalent colonization factors. It is possible that there are additional lineages that are yet to be defined among these less frequent isolates or that these isolates represent ETEC that have recently acquired virulence plasmids.

Our data clearly demonstrate that ETEC harbor identifiable lineages, with the majority containing consistent, definable virulence profiles. This finding implies that, in these lineages, the virulence determinants were acquired once and the clades subsequently spread,

© 2014 Nature America, Inc. All rights reserved. npg

in some instances, around the world. This narrative is markedly different from the one that could be expected on the basis of previous work on this pathovar, and it is clear from our analysis that the ETEC population contains a number of stable clones with specific virulence profiles. Similar patterns have been identified in other pathogens. Detailed analyses of the five major lineages (L1–L5) using Bayesian-based analysis suggest that these lineages emerged between 51 and 174 years ago (**Fig. 2**). This observation is consistent with the time period of the spread of several other major pathogens, including *Vibrio cholerae*[28], *Staphylococcus aureus*[29], *Shigella sonnei*[30] and invasive *Salmonella* Typhimurium[31]. The precise cause of the emergence and spread of these organisms is likely multifactorial, but factors such as international travel have certainly contributed to the spread of these pathogens around the world.

In summary, the data presented here show that ETEC-mediated disease is actually a set of overlapping global epidemics of individual ETEC lineages, which have been stable over substantial periods of time in endemic areas. Furthermore, these globally distributed lineages with consistent colonization factor profiles seem to cause disease in children and adults in endemic areas and travelers to the same extent. It is of particular interest that our data suggest that plasmid acquisition is a major vehicle driving the emergence of different ETEC clades. This suggests that the development of a vaccine based on the most prevalent colonization factors could be protective against a large proportion of ETEC diarrhea cases. New ETEC vaccines are under development that are based on the major colonization factors identified in ETEC causing diarrhea[5,32]. Hence, our study has key implications for how ETEC disease is understood and tracked, and may help in disease prevention.

URLs. Source code for SNP calling, https://github.com/sanger-pathogens/snp_sites.

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. Primary accession codes for the Illumina sequence reads of all 362 ETEC isolates sequenced in this study are included in **Supplementary Table 3.**

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

This work was supported by the Wellcome Trust (grant 098051), the Swedish Research Council (grants 2012-3464 and 2011-3435), the Swedish Strategic Foundation (grant SB12-0072) and the European Research Council (grant 239784).

AUTHOR CONTRIBUTIONS

A.v.M., G.D., A.-M.S., Å.S., T.R.C. and D.A.R. contributed to the design of the study and data interpretation. A.v.M. and G.W. extracted DNA. A.v.M. screened the sequence data and performed the majority of the bioinformatics analyses with input from T.R.C. and N.R.T. A.v.M. interpreted and analyzed the results from the recombination detection and BAPS analyses, executed by J.C. T.S. and L.H.W. identified the MCG and determined sequence types from whole-genome data. A.I. performed the BLASTN analysis to identify O antigen genotypes in all ETEC isolates included. E.J. analyzed the genes encoding toxins. D.P. was responsible for forwarding extracted DNA samples to the sequencing pipeline at the Wellcome Trust Sanger Institute. G.D., Å.S. and A.-M.S. supervised the work. All authors contributed to the writing of the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. World Health Organization. *Diarrhoeal Diseases* (*Updated February 2009*) (World Health Organization, Geneva, 2009).
2. Gaastra, W. & Svennerholm, A.M. Colonization factors of human enterotoxigenic *Escherichia coli* (ETEC). *Trends Microbiol.* **4**, 444–452 (1996).
3. Qadri, F., Svennerholm, A.M., Faruque, A.S. & Sack, R.B. Enterotoxigenic *Escherichia coli* in developing countries: epidemiology, microbiology, clinical features, treatment, and prevention. *Clin. Microbiol. Rev.* **18**, 465–483 (2005).
4. Isidean, S.D., Riddle, M.S., Savarino, S.J. & Porter, C.K. A systematic review of ETEC epidemiology focusing on colonization factor and toxin expression. *Vaccine* **29**, 6167–6178 (2011).
5. Svennerholm, A.M. & Lundgren, A. Recent progress toward an enterotoxigenic *Escherichia coli* vaccine. *Expert. Rev. Vaccines* **11**, 495–507 (2012).
6. Wolf, M.K. Occurrence, distribution, and associations of O and H serogroups, colonization factor antigens, and toxins of enterotoxigenic *Escherichia coli. Clin. Microbiol. Rev.* **10**, 569–584 (1997).
7. Smith, H.W. The exploitation of transmissible plasmids to study the pathogenesis of *E. coli* diarrhoea. *Proc. R. Soc. Med.* **66**, 272–273 (1973).
8. Turner, S.M. *et al.* Phylogenetic comparisons reveal multiple acquisitions of the toxin genes by enterotoxigenic *Escherichia coli* strains of different evolutionary lineages. *J. Clin. Microbiol.* **44**, 4528–4536 (2006).
9. Steinsland, H., Lacher, D.W., Sommerfelt, H. & Whittam, T.S. Ancestral lineages of human enterotoxigenic *Escherichia coli. J. Clin. Microbiol.* **48**, 2916–2924 (2010).
10. Escobar-Páramo, P. *et al.* A specific genetic background is required for acquisition and expression of virulence factors in *Escherichia coli. Mol. Biol. Evol.* **21**, 1085–1094 (2004).
11. Regua-Mangia, A.H. *et al.* Genotypic and phenotypic characterization of enterotoxigenic *Escherichia coli* (ETEC) strains isolated in Rio de Janeiro city, Brazil. *FEMS Immunol. Med. Microbiol.* **40**, 155–162 (2004).
12. Steinsland, H., Valentiner-Branth, P., Aaby, P., Mølbak, K. & Sommerfelt, H. Clonal relatedness of enterotoxigenic *Escherichia coli* strains isolated from a cohort of young children in Guinea-Bissau. *J. Clin. Microbiol.* **42**, 3100–3107 (2004).
13. Valvatne, H., Steinsland, H. & Sommerfelt, H. Clonal clustering and colonization factors among thermolabile and porcine thermostable enterotoxin-producing *Escherichia coli. APMIS* **110**, 665–672 (2002).
14. Sahl, J.W. & Rasko, D.A. Analysis of global transcriptional profiles of enterotoxigenic *Escherichia coli* isolate E24377A. *Infect. Immun.* **80**, 1232–1242 (2012).
15. Herzer, P.J., Inouye, S., Inouye, M. & Whittam, T.S. Phylogenetic distribution of branched RNA-linked multicopy single-stranded DNA among natural isolates of *Escherichia coli. J. Bacteriol.* **172**, 6175–6181 (1990).
16. Corander, J., Marttinen, P., Sirén, J. & Tang, J. Enhanced Bayesian modelling in BAPS software for learning genetic structures of populations. *BMC Bioinformatics* **9**, 539 (2008).
17. Lasaro, M.A., Mathias-Santos, C., Rodrigues, J.F. & Ferreira, L.C.S. Functional and immunological characterization of a natural polymorphic variant of a heat-labile toxin (LT-I) produced by enterotoxigenic *Escherichia coli* (ETEC). *FEMS Immunol. Med. Microbiol.* **55**, 93–99 (2009).
18. Rodrigues, J. *et al.* Clonal structure and virulence factors in strains of *Escherichia coli* of the classic serogroup O55. *Infect. Immun.* **64**, 2680–2686 (1996).
19. Froehlich, B., Parkhill, J., Sanders, M., Quail, M.A. & Scott, J.R. The pCoo plasmid of enterotoxigenic *Escherichia coli* is a mosaic cointegrate. *J. Bacteriol.* **187**, 6509–6516 (2005).
20. Johnson, T.J. & Nolan, L.K. Pathogenomics of the virulence plasmids of *Escherichia coli. Microbiol. Mol. Biol. Rev.* **73**, 750–774 (2009).
21. Crossman, L.C. *et al.* A commensal gone bad: complete genome sequence of the prototypical enterotoxigenic *Escherichia coli* strain H10407. *J. Bacteriol.* **192**, 5822–5831 (2010).
22. Croucher, N.J. *et al.* Rapid pneumococcal evolution in response to clinical interventions. *Science* **331**, 430–434 (2011).
23. Didelot, X. *et al.* Microevolutionary analysis of *Clostridium difficile* genomes to investigate transmission. *Genome Biol.* **13**, R118 (2012).
24. Fischer Walker, C.L., Perin, J., Aryee, M.J., Boschi-Pinto, C. & Black, R.E. Diarrhea incidence in low- and middle-income countries in 1990 and 2010: a systematic review. *BMC Public Health* **12**, 220 (2012).
25. Sahl, J.W. *et al.* A comparative genomic analysis of diverse clonal types of enterotoxigenic *Escherichia coli* reveals pathovar-specific conservation. *Infect. Immun.* **79**, 950–960 (2011).
26. Sabui, S. *et al.* Allelic variation in colonization factor CS6 of enterotoxigenic *Escherichia coli* isolated from patients with acute diarrhoea and controls. *J. Med. Microbiol.* **59**, 770–779 (2010).
27. Pupo, G.M., Karaolis, D.K., Lan, R. & Reeves, P.R. Evolutionary relationships among pathogenic and nonpathogenic *Escherichia coli* strains inferred from multilocus enzyme electrophoresis and mdh sequence studies. *Infect. Immun.* **65**, 2685–2692 (1997).
28. Mutreja, A. *et al.* Evidence for several waves of global transmission in the seventh cholera pandemic. *Nature* **477**, 462–465 (2011).
29. Harris, S.R. *et al.* Evolution of MRSA during hospital transmission and intercontinental spread. *Science* **327**, 469–474 (2010).
30. Holt, K.E. *et al. Shigella sonnei* genome sequencing and phylogenetic analysis indicate recent global dissemination from Europe. *Nat. Genet.* **44**, 1056–1059 (2012).
31. Okoro, C.K. *et al.* Intracontinental spread of human invasive *Salmonella* Typhimurium pathovariants in sub-Saharan Africa. *Nat. Genet.* **44**, 1215–1221 (2012).
32. Darsley, M.J. *et al.* The oral, live attenuated enterotoxigenic *Escherichia coli* vaccine ACE527 reduces the incidence and severity of diarrhea in a human challenge model of diarrheal disease. *Clin. Vaccine Immunol.* **19**, 1921–1931 (2012).

© 2014 Nature America, Inc. All rights reserved.

ONLINE METHODS

Isolate selection. In total, 362 human ETEC isolates from the Gothenburg University (UG) ETEC collection, comprising more than 3,500 ETEC isolates, were selected on the basis of virulence factor profile, origin and year of isolation to represent a broad collection of ETEC isolated worldwide. This collection included attempts to cover the most prevalent colonization factor and toxin profiles as well as isolates with rare colonization factors or lacking identifiable colonization factors. In addition, isolates were from as many geographical locations as possible collected over a long time period. Isolates from different patient groups were also included, i.e., from children and adults in ETEC-endemic areas as well as from travelers and soldiers visiting such areas. Most isolates were from subjects suffering from diarrhea, both hospitalized and treated as outpatients; some isolates were derived from asymptomatic carriers. The isolates were from several countries in Asia, Africa, and North, Central and South America collected between 1980 and 2011 (**Supplementary Table 2**). ETEC isolates were identified by culture on MacConkey agar followed by analysis of LT and ST toxin expression using GM1 ELISAs and in some cases PCR[33]. Different colonization factors were identified by dot-blot analyses and in some cases by multiplex PCR[34]. Isolates had been kept in glycerol stocks at −70 °C, and each isolate had been passaged as few times as possible. The strains were collected with informed consent from patients or the parents of children. Permission to use the ETEC strain collection was granted by the Regional Ethical Board of Gothenburg, Sweden (Ethics Committee Reference 088-10).

Bacterial culture and DNA preparation. All isolates were initially cultured on horse blood agar plates overnight at 37 °C to detect possible contamination. Pure ETEC cultures were used for DNA extraction with the Wizard Genomic DNA kit (Promega) according to the manufacturer's instructions. DNA quantity was measured by NanoDrop spectrophotometer (NanoDrop Technologies). A DNA concentration of at least 72 ng/μl was used for each isolate in Illumina sequencing.

Genomic library preparation and DNA sequencing. Paired-end sequencing was performed on the Illumina HiSeq 2000 platform. Libraries were constructed, according to the protocols of Quail *et al.*[35], with a fragment length of 75 or 100 bases in pools of uniquely tagged isolates. In total, 362 isolates were sequenced with a read length of 75 or 100 bases. The difference in read length did not interfere with assembly. Index tagging sequences were used to assign reads to individual samples for further analysis.

***De novo* sequence assembly.** Paired-end Illumina sequence data from each isolate were *de novo* assembled using the Velvet pipeline developed in house at the Wellcome Trust Sanger Institute. The pipeline consisted of the following steps. First, Velvet Optimiser fragmented the reads into shorter sequences (66–90% of original read length), the best parameter set was chosen on the basis of N50 values and Velvet was run to produce a set of contigs. Second, contigs of less than 300 bases in length were removed. Third, the scaffolding software SSPACE (v 2.0) was run to scaffold the assembly from the previous step. Fourth, gaps were filled by GapFiller (v 1.10). Fifth, reads were mapped to the scaffolds from the third step, and statistics and graphs were generated for this assembly. In total, four samples were removed from the data set because of contamination or incomplete assembly. The best assembly for each isolate was chosen on the basis of a combination of contig length, contig number and N50 value.

Determination of the maximum common genome and SNP detection. The genomes of 47 fully sequenced *E. coli* strains available in GenBank (June 2012) were used. We extracted a set of 1,429 genes that were termed the maximum common genome (MCG), i.e., all genes that were present in each of the 47 genomes. Genes with at least 70% identity on the protein level were considered to form the MCG. This was determined by hierarchical clustering using CD-HIT[36]. The protein sequences of the core genes were concatenated for each of the 47 genomes, and a maximum-parsimony phylogeny was created. Extracted core genes for all 362 ETEC genomes were compared, and SNPs were detected using an in-house script (source code can be found on GitHub; see URLs).

Recombination analysis. The BratNextGen method[37] was used to detect recombination events with the concatenated MCG alignment for 362 ETEC isolates and 21 *E. coli* references. Estimation of recombination was performed with the default settings as in refs. 37–39, using 20 iterations of the estimation algorithm, which was assessed to be sufficient as changes in the hidden Markov model parameters were already negligible over the last 70% of the iterations. The significance of a recombining region was determined as in ref. 37 using a permutation test with 100 permutations executed in parallel on a cluster computer with the threshold of 5% to conclude significance for each region. Every significant recombination was then masked as missing data in the MCG alignment to provide robust input data for phylogenetic and population genetic analyses.

Phylogenetic analysis. In total, 1,429 genes constituting the MCG of 47 well-annotated *E. coli* and *Shigella* strains were extracted. All phylogenetic trees were generated by FastTree (v2.1.4) for all variable sites using a general time-reversible model with gamma correction for among-site rate variation for ten initial trees and the maximum-likelihood algorithm. Phylogenetic trees are based on the alignment of the 1,429 genes, with recombination sites masked. In addition to the 362 sequenced ETEC isolates, 21 *E. coli* reference strains were included (**Supplementary Table 1**).

Population genetic analysis. To estimate the population structure, we used the BAPS v6.0 software[16,40], in particular, its module hierBAPS[41], which fits lineages to genome data using nested clustering. BAPS has been shown to efficiently estimate bacterial population structure from both limited core genome variation[42–44] and whole-genome sequence data[28,38,39,45]. Two nested levels of molecular variation were fitted to the MCG alignment with estimated recombinations masked as missing data. The estimation used 10 independent runs of the stochastic optimization algorithm with the a priori upper bound of the number of clusters varying over the interval 50–150 across the runs. The estimated mode of the posterior distribution had 8 and 39 clusters at levels 1 and 2 of the hierarchy, respectively. All clusters were significantly supported when compared against alternative partitions (posterior probability for any cluster of at least 100-fold higher than for the alternative).

Bayesian phylogeny and estimating dates of emergence of lineages. Estimation of rates of evolution and the age of individual clusters was performed using Bayesian inference, BEAST (v5.8.8)[46], on SNP alignments. Various combinations of a population size change model and molecular clock model were compared to find the model best fitting the data. In all cases, the Bayes factor showed strong support (Bayes factor >> 200) for the use of a skyline[47] model of population size change and a relaxed uncorrelated lognormal clock, which allows evolutionary rates to change among the branches of the tree[28], and a Hasegawa, Kishino and Yano (HKY) substitution model with gamma-distributed rate heterogeneity among sites was used[48]. In all cases, three independent chains were run for 500 million steps each with sampling every 10,000 steps. The three chains were combined with LogCombiner[46], with the first 50 million steps removed from each as a burn-in. MCC trees were created and annotated using TreeAnnotator and were viewed in FigTree[46]. We report estimates as median values within 95% highest probability density (HPD) intervals and report posterior probability values as support for identified ancestral node age.

Sequence-based determination of O antigens. On the basis of a sequence data set for two O antigen–processing gene sets (*wzx*/*wzy* and *wzm*/*wzt*) for all 184 *E. coli* O antigens (A.I., S. Iyoda, T. Kikuchi, Y. Ogura and K. Katsura, unpublished data) and the results of BLASTN analysis, we identified the ETEC O antigen genotype (defined by ≥97% sequence identity and ≥97% aligned length coverage of a query sequence, in both genes from a set) from the assembled draft genomes. Moderately diversified sequences (defined by 70–97% sequence identity and ≥70% aligned length coverage) were classified into '-like' categories. For isolates with previously unknown O antigen genotypes, *wzx*/*wzy* and *wzm*/*wzt* sequences were extracted from the regions of the draft genomes encoding O antigen biosynthesis genes and classified into groups that were categorized as novel *E. coli* O antigen genotypes (defined by ≥97% sequence identity and ≥97% aligned length coverage).

© 2014 Nature America, Inc. All rights reserved.

Screening for plasmid incompatibility groups. In total, 362 ETEC isolates were screened for known plasmid incompatibility groups using *in silico* PCR with primers for the following incompatibility groups: FII, FIIK, FIIS, FIIY, FIA, BIB, FrepB, HI1, I1, K, N, P, X1, X4 and Y[49].

33. Sjöling, A., Wiklund, G., Savarino, S.J., Cohen, D.I. & Svennerholm, A.M. Comparative analyses of phenotypic and genotypic methods for detection of enterotoxigenic *Escherichia coli* toxins and colonization factors. *J. Clin. Microbiol.* **45**, 3295–3301 (2007).
34. Rodas, C. *et al.* Development of multiplex PCR assays for detection of enterotoxigenic *Escherichia coli* colonization factors and toxins. *J. Clin. Microbiol.* **47**, 1218–1220 (2009).
35. Quail, M.A., Swerdlow, H. & Turner, D.J. Improved protocols for the Illumina Genome Analyzer sequencing system. *Curr. Protoc. Hum. Genet.* **Unit 18.2** 1–27 (2009).
36. Li, W. & Godzik, A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**, 1658–1659 (2006).
37. Marttinen, P. *et al.* Detection of recombination events in bacterial genomes from large population samples. *Nucleic Acids Res.* **40**, e6 (2012).
38. McNally, A., Cheng, L., Harris, S.R. & Corander, J. The evolutionary path to extraintestinal pathogenic, drug-resistant *Escherichia coli* is marked by drastic reduction in detectable recombination within the core genome. *Genome Biol. Evol.* **5**, 699–710 (2013).
39. Castillo-Ramírez, S. *et al.* Phylogeographic variation in recombination rates within a global clone of methicillin-resistant *Staphylococcus aureus*. *Genome Biol.* **13**, R126 (2012).
40. Tang, J., Hanage, W.P., Fraser, C. & Corander, J. Identifying currents in the gene pool for bacterial populations using an integrative approach. *PLOS Comput. Biol.* **5**, e1000455 (2009).
41. Cheng, L., Connor, T.R., Sirén, J., Aanensen, D.M. & Corander, J. Hierarchical and spatially explicit clustering of DNA sequences with BAPS software. *Mol. Biol. Evol.* **30**, 1224–1228 (2013).
42. Corander, J., Connor, T.R., O'Dwyer, C.A., Kroll, J.S. & Hanage, W.P. Population structure in the *Neisseria*, and the biological significance of fuzzy species. *J. R. Soc. Interface* **9**, 1208–1215 (2012).
43. Willems, R.J.L. *et al.* Restricted gene flow among hospital subpopulations of *Enterococcus faecium*. *MBio* **3**, e00151-12 (2012).
44. Hanage, W.P., Fraser, C., Tang, J., Connor, T.R. & Corander, J. Hyper-recombination, diversity, and antibiotic resistance in pneumococcus. *Science* **324**, 1454–1457 (2009).
45. Croucher, N.J. *et al.* Population genomics of post-vaccine changes in pneumococcal epidemiology. *Nat. Genet.* **45**, 656–663 (2013).
46. Drummond, A.J. & Rambaut, A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* **7**, 214 (2007).
47. Drummond, A.J., Rambaut, A., Shapiro, B. & Pybus, O.G. Bayesian coalescent inference of past population dynamics from molecular sequences. *Mol. Biol. Evol.* **22**, 1185–1192 (2005).
48. Drummond, A.J., Ho, S.Y., Phillips, M.J. & Rambaut, A. Relaxed phylogenetics and dating with confidence. *PLoS Biol.* **4**, e88 (2006).
49. Carattoli, A. *et al.* Identification of plasmids by PCR-based replicon typing. *J. Microbiol. Methods* **63**, 219–228 (2005).

© 2014 Nature America, Inc. All rights reserved.

doi:10.1038/ng.3145

Journal of Food Protection, Vol. 77, No. 8, 2014, Pages 1269–1274
doi:10.4315/0362-028X.JFP-13-506
Copyright ©, International Association for Food Protection

Identification of O Serotypes, Genotypes, and Virulotypes of Shiga Toxin–Producing *Escherichia coli* Isolates, Including Non-O157 from Beef Cattle in Japan

HIROHISA MEKATA,[1] ATSUSHI IGUCHI,[2] KIMIKO KAWANO,[3] YUMI KIRINO,[1] IKUO KOBAYASHI,[4] AND NAOAKI MISAWA[3*]

[1]*Project for Zoonoses Education and Research, Faculty of Agriculture, and* [3]*Center for Animal Disease Control, University of Miyazaki, 1-1 Gakuen-Kibanadai-Nishi, Miyazaki, 889-2192, Japan;* [2]*Interdisciplinary Research Organization, University of Miyazaki, 5200 Kiyotake, Miyazaki, 889-1692, Japan; and* [4]*Sumiyoshi Livestock Science Station, Field Science Center, Faculty of Agriculture, University of Miyazaki, 10100-1 Shimanouchi, Miyazaki 880-0121, Japan*

MS 13-506: Received 25 November 2013/Accepted 3 April 2014

ABSTRACT

Bovines are recognized as an important reservoir of Shiga toxin–producing *Escherichia coli* (STEC). Although STEC strains are significant foodborne pathogens, not all of the STEC held by cattle are pathogenic, and which type of STEC that will become epidemic in humans is unpredictable. Information about the prevalence of serotype and virulence gene distribution in beef cattle is insufficient to develop monitoring and controlling activities for a food safety and security program. Thus, this study investigated the prevalence of O157 and non-O157 STEC in Japanese beef cattle and characterized the isolates by the type of O antigen and several virulence markers to help predict the pathogenicity. In this study, 64.2% (176 of 274) of enrichment cultures of fecal samples collected from an abattoir and farms were stx_1 and/or stx_2 positive by PCR. STEC strains were isolated from 22.1% (39 of 176) of the positive fecal samples, and these isolates represented 17 types of O antigen (O1, O2 or O50, O5, O8, O55, O84, O91, O109, O113, O136, O150, O156, O157, O163, O168, O174, and O177). Two selective media targeting major STEC groups, cefixime-tellurite sorbitol MacConkey agar and CHROMagar O26/O157, allowed isolation of a variety of STEC strains. The most frequently isolated STEC was O113 (8 of 39), which has previously been reported as a cause of foodborne infections. Although most of the O113 STEC isolated from infected patients possessed the enterohemolysin (*hlyA*) gene, none of the O113 STEC cattle isolates possessed the *hlyA* gene. The second most common isolate was O157 (6 of 39), and all these isolates contained common virulence factors, including *eae*, *tir*, lpf_1, lpf_2, and *hlyA*. This study shows the prevalence of O157 and non-O157 STEC in Japanese beef cattle and the relationship of O antigen and virulotypes of the isolates. This information may improve identification of the source of infection, developing surveillance programs or the current understanding of virulence factors of STEC infections.

Shiga toxin–producing *Escherichia coli* (STEC), alternatively referred to as verocytotoxigenic *E. coli*, is a significant foodborne pathogen causing diarrhea, hemorrhagic colitis, and hemolytic uremic syndrome *(17, 20)*. More than 100 types of O antigens from STEC isolates have thus far been identified, and O157 STEC has been the main serotype causing STEC outbreaks in the world *(9, 21, 38)*. Recently, several non-O157 STEC have also been identified as the cause of severe illnesses *(6, 19)*. In Japan, infections caused by non-O157 serotypes have increased annually and accounted for 47% of all STEC infections in 2012 *(28)*.

The pathogenicity of STEC in humans depends on a number of virulence factors, such as capacity to produce the Shiga toxin and attach to target cells *(13, 24)*. Shiga toxin is the most important virulence factor in STEC and can be differentiated into two major groups: stx_1 and stx_2. Several variants (stx_{1a}, stx_{1c}, stx_{1d}, and stx_{2a} to stx_{2g}) have been identified, and a close association has been found between the severity of disease and the presence of the variants, such as stx_{2a} and/or stx_{2c} *(4, 14, 30)*. In addition to these, STEC strains have a number of other virulence factors. One of the major virulence factors, intimin, is associated with attachment to the host's intestinal mucosa conferred by the gene *eaeA*, and its receptor is encoded by the *tir* gene *(35, 44, 46)*. Other virulence genes associated with attachment to cells include the long polar fimbriae (*lpf*) and the STEC auto-agglutinating adhesion (*saa*) gene *(5, 43)*. Additionally, the enterohemolysin (*hlyA*) and subtilase (*subAB*) virulence factors were found to associate with hemorrhagic enterocolitis *(33, 39)*.

Bovines are recognized as a primary source of STEC, and many outbreaks are linked to bovine food sources *(1, 3, 12, 23)*. In addition, consumption of water or vegetables contaminated with bovine feces is frequently involved in outbreaks *(18, 29)*. Despite an increase in the number of STEC outbreaks, there is a lack of data on prevalence at the farm level, distribution of serotypes, and virulence factors. To

* Author for correspondence. Tel and Fax: (81)-985-58-7284; E-mail: a0d901u@cc.miyazaki-u.ac.jp.

TABLE 1. *Target and primer sequences used for the detection of STEC virulence genes*

Target gene	Primer	Primer sequence	Reference
stx_1	stx1F	5′-ATAAATCGCCATTCGYTGACTAC-3′	*31*
	stx1R	5′-AGAACGYCCACTGAGATCATC-3′	
stx_2	stx2F	5′-GGCACTGTCTGAAACTGCTCC-3′	*31*
	stx2R	5′-TCGCCASTTATCTGACATTCTG-3′	
eae	eaeAF	5′-GACCCGGCACAAGCATAAGC-3′	*31*
	eaeAR	5′-CCACCTGCAGCAACAAGAGG-3′	
vtx_{1a}	vtx1a-F1	5′-CCTTTCCAGGTACAACAGCGGTT-3′	*45*
	vtx1a-R2	5′-GGAAACTCATCAGATGCCATTCTGG-3′	
vtx_{1c}	vtx1c-F1	5′-CCTTTCCTGGTACAACTGCGGTT-3′	*45*
	vtx1c-R1	5′-CAAGTGTTGTACGAAATCCCCTCTGA-3′	
vtx_{1d}	vtx1d-F1	5′-CAGTTAATGCGATTGCTAAGGAGTTTACC-3′	*45*
	vtx1d-R1	5′-CTCTTCCTCTGGTTCTAACCCCATGATA-3′	
vtx_{2a}	vtx2a-F2	5′-GCGATACTGRGBACTGTGGCC-3′	*45*
	vtx2a-R2	5′-GGCCACCTTCACTGTGAATGTG-3′	
	vtx2a-R3	5′-CCGKCAACCTTCACTGTAAATGTG-3′	
vtx_{2b}	vtx2b-F1	5′-AAATATGAAGAAGATATTTGTAGCGGC-3′	*45*
	vtx2b-R1	5′-CAGCAAATCCTGAACCTGACG-3′	
vtx_{2c}	vtx2c-F1	5′-GAAAGTCACAGTTTTTATATACAACGGGTA-3′	*45*
	vtx2c-R2	5′-CCGGCCACYTTTACTGTGAATGTA-3′	
vtx_{2d}	vtx2d-F1	5′-AAARTCACAGTCTTTATATACAACGGGTG-3′	*45*
	vtx2d-R1	5′-TTYCCGGCCACTTTTACTGTG-3′	
	vtx2d-R2	5′-GCCTGATGCACAGGTACTGGAC-3′	
vtx_{2e}	vtx2e-F1	5′-CGGAGTATCGGGGAGAGGC-3′	*45*
	vtx2e-R2	5′-CTTCCTGACACCTTCACAGTAAAGGT-3′	
vtx_{2f}	vtx2f-F1	5′-TGGGCGTCATTCACTGGTTG-3′	*45*
	vtx2f-R1	5′-TAATGGCCGCCCTGTCTCC-3′	
vtx_{2g}	vtx2g-F1	5′-CACCGGGTAGTTATATTTCTGTGGATATC-3′	*45*
	vtx2g-R1	5′-GATGGCAATTCAGAATAACCGCT-3′	
hlyA	hlyAF	5′-GCATCATCAAGCGTACGTTCC-3′	*31*
	hlyAR	5′-AATGAGCCAAGCTGGTTAAGCT-3′	
saa	SAADF	5′-CGTGATGAACAGGCTATTGC-3′	*32*
	SAADR	5′-ATGGACATGCCTGTGGCAAC-3′	
tir	TIR-F	5′-CATTACCTTCACAAACCGAC-3′	*22*
	TIR-R	5′-CCCCGTTAATCCTCCCAT-3′	
lpf_1	lpfO141-F	5′-CTGCGCATTGCCGTAAC-3′	*41*
	lpfO141-R	5′-ATTTACAGGCGAGATCGTG-3′	
lpf_2	lpfA-F	5′-ATGAAGCGTAATATTATAG-3′	*11*
	lpfA-R	5′-TTATTTCTTATATTCGAC-3′	
subAB	SubAF	5′-GTACGGACTAACAGGGAACTG-3′	*33*
	SubAR	5′-ATCGTCATATGCACCTCCG-3′	

establish monitoring programs and control methods of STEC contamination in food or the environment, we need to understand the prevalence of STEC, especially non-O157 STEC in cattle and its pathogenesis in humans. Therefore, this study investigated the relationship between the types of O antigen and the presence of virulence genes among the isolates.

MATERIALS AND METHODS

Sample collection. A total of 274 fresh cattle fecal samples were obtained from June 2012 to February 2013. Of these 274 cattle, 140 were from an abattoir, and the other 134 were from five individual farms located in different areas. Most samples were removed from animals, and some were obtained after defecation with attention to contamination. All the samples were transported to the laboratory within 1 h and stored at 4°C for no more than a 24 h prior to analysis.

Bacterial cultures and detection of *stx*-positive *E. coli*. Each fecal sample (1 g) was mixed by vortexing in 10 ml of modified *E. coli* broth (Nissui, Tokyo, Japan) and incubated overnight at 42°C. A 100-μl aliquot of each incubated broth was directly boiled for 10 min to extract the genomic DNA and centrifuged for 3 min (10,000 × *g* at 4°C). The supernatant was used as template DNA, and the remainder of the enrichment samples were stored at 4°C until the result of *stx* screening was obtained. The extracted DNAs were screened for the presence of stx_1 and/or stx_2 genes by PCR using a thermal cycler (Veriti 96-Well Thermal Cycler, Applied Biosystems, Foster City, CA). The primer sets and target genes are listed in Table 1 (*11, 22, 31–33, 41, 45*). To neutralize the PCR inhibitor included in the template DNA, Ampdirect Plus (Shimadzu Biotech Co., Tsukuba, Japan) was used for the PCR. The PCR amplification was carried out in a reaction mixture containing 10 μl of 2 × Ampdirect Plus, 1.0 μl of primer (100 pmol each), 0.1 μl of BIOTAQ HS DNA Polymerase (Bioline, London, UK), 1.0 μl of template DNA sample, and PCR-grade water to increase the volume to 20 μl. The reaction mixture was the same for both primer sets. Positive and blank control samples were included in each set of reactions. The cycling

TABLE 2. *The types and number of agar plates used for isolation of each O type STEC in this study*

Type of O antigen	No. of isolates	DHL	CT-sMAC	CHROMagar O26/O157
O1B	5	3	4	4
O2 or O50[a]	1	1	0	0
O5	1	0	0	1
O8	1	0	0	1
O55	1	0	1	0
O84	1	1	1	1
O91	1	1	1	0
O109	5	2	2	3
O113	8	1	8	5
O136	2	0	2	1
O150	1	0	0	1
O156	1	0	1	1
O157	6	1	5	6
O163	1	1	1	0
O168	1	0	1	1
O174	1	0	1	1
O177	1	0	0	1
OUT	1	0	1	1
Total	39	11	29	28

[a] O2 or O50: indistinctive.

conditions had an initial denaturation at 95°C for 10 min, followed by 40 cycles of 95°C for 30 s, 60°C for 1 min, and 72°C for 1 min. A final extension was performed at 72°C for 7 min. The PCR products were resolved by electrophoresis through a 2% agarose gel and were visualized under UV light (AE-6932GXCF, ATTO Corporation, Tokyo, Japan) by GelRed Nucleic Acid Gel Stain (Biotium, Hayward, CA). The product size was estimated using a 100-bp DNA ladder (TaKaRa Bio Company, Otsu, Japan).

Isolation of *stx*-positive *E. coli*. Then, 500 μl of each sample of stx_1 or/and stx_2 PCR-positive enrichment culture was centrifuged (10,000 × *g*, 5 min at 4°C). Because hydrochloric acid treatment is effective for isolating STEC from other bacteria, the resultant pellets were diluted in 100 μl of saline and treated with 100 μl of 0.5% NaCl and 1/8 N HCl for 30 s *(15, 16)*. Finally, these were plated onto desoxycholate hydrogen sulfide lactose (DHL) agar (Nissui), CHROMagar O26/O157 (CHROMagar Microbiology, Paris, France), and sorbitol MacConkey agar (Nissui) supplemented with cefixime-tellurite supplement (CT-sMAC, Merck, Darmstadt, Germany). After overnight incubation at 37°C, 4 to 6 colonies with differing colony color or morphology were taken from each plate and screened for the presence of *stx* genes by PCR amplification, as described above. *stx*-positive colonies were streaked onto Luria-Bertani (LB) agar (Nacalai Tesque, Kyoto, Japan) and incubated overnight at 37°C. Each streaked STEC isolate was individually cultured in 2 ml of LB broth (Nacalai Tesque) overnight at 37°C. Genomic DNA was extracted from a cell pellet by boiling for 10 min. All STEC isolates were preserved in LB broth with 10% glycerol at −80°C and were routinely recultured during the study.

Serotyping and genotyping. Serotypes of the O antigen were identified by the 50 types of antisera (Denka Seiken, Tokyo, Japan). Genotypes of O antigen were identified by multiplex PCR, which detects the specific sequence (mostly *wzx, wzy, wzt,* and *wzm* genes) of each O type of *E. coli*. This typing method was developed by Dr. Iguchi Interdisciplinary Research Organization at the University of Miyazaki (available at: http://www.cc.miyazaki-u.ac.jp/iguchi/iguchi_lab/O-genotyping.html).

Detection of virulence-related genes. The template DNA from each STEC isolate was tested by PCR for the presence of a range of virulence genes, including *eae, hlyA, saa, tir*, lpf_1, lpf_2, and *subAB* (Table 1). In the initial analysis, sequence analysis using nucleotide BLAST was performed to confirm the specificity of the PCR products. After this analysis, each PCR product size was checked by electrophoresis, and the specificity of the analyses was confirmed.

Determination of *stx* variants. Each STEC Shiga toxin gene was subtyped by PCR, according to the previous report *(45)*. The annealing temperatures and PCR conditions were slightly modified for a thermal cycler and *Taq* polymerase. Briefly, annealing temperatures of stx_1 and stx_2 variants were 66°C, except for vtx_{2c} (69°C) and vtx_{2d} (69°C).

RESULTS

Of the 274 fecal samples analyzed, 176 (64.2%) were stx_1 and/or stx_2 positive, and stx_1, stx_2, and the combination of both were detected from 42 (15.3%), 54 (19.7%), and 80 (29.1%) samples, respectively. STEC strains were isolated from 39 (22.1%) *stx*-positive samples. Of these STEC samples, 11, 29, and 28 isolates were isolated from DHL, CT-sMAC, and CHROMagar O26/O157 plates, respectively (Table 2). By the O-genotyping methods, 38 isolates were obtained representing 17 types of O antigen (O1, O2 or O50, O5, O8, O55, O84, O91, O109, O113, O136, O150, O156, O157, O163, O168, O174, and O177), and only one isolate could not be assigned (Table 3). On the other hand, 16 isolates were typeable (O1, O55, O91, O136, O157, and O168) and 23 were not typeable by the O serotyping method because only 50 types of O antisera were available. The most frequently occurring types of O antigen were O113 (20.5%), O157 (15.3%), O109 (12.8%), and O1 (12.8%).

The stx_{1d}, stx_{2b}, stx_{2e}, and stx_{2f} gene variants were not detected; however, stx_{1a}, stx_{1c}, stx_{2a}, stx_{2c}, stx_{2d}, and stx_{2g} were present in 15 (38.4%), 7 (17.9%), 18 (46.1%), 8 (20.5%), 8 (20.5%), and 1 (2.5%) of the isolates, respectively. Interestingly, the same *stx* gene variants were detected within strains from the same type of O antigen (Table 3). For example, all O1 and O109 strains possessed vtx_{2a}. Likewise, all O113, O136, and O157 isolates possessed vtx_{2d}, vtx_{1a}, and vtx_{2c}, respectively. The intimin gene (*eaeA*) was present in 38.4% (15 of 39) of isolates: serotypes O5, O84, O109, O150, O156, O157, and O177. The enterohemolysin gene (*hlyA*) was present in 64.1% (25 of 39). Other virulence-related genes, including *saa, tir*, lpf_1, lpf_2, and *subAB* were detected. Although O157 STEC strains possessed common virulence-related genes, non-O157 STEC tended to possess different profiles of virulence-related genes.

DISCUSSION

To prescreen samples before subjecting them to cultural procedures, PCR analyses, generally targeting the *stx* genes, are employed in many of the livestock fecal prevalence studies *(2, 8, 42)*. This study detected the *stx* genes in 64.2% of cattle feces. Previous studies have reported STEC

TABLE 3. *O antigen,* stx *variants, and virulence markers of STEC isolated from Japanese beef cattle*[a]

Genotype	Serotype[b]	No. of isolates (abattoir, farm)	*vtx*$_1$		*vtx*$_2$				*eae*	*tir*	*saa*	*lpf*$_1$	*lpf*$_2$	*hlyA*	*subAB*
			a	c	a	c	d	g							
O1B	O1	3 (2, 1)	−	−	+	−	−	−	−	−	−	−	−	+	−
O1B	O1	2 (1, 1)	−	−	+	−	−	−	−	−	−	−	−	−	−
O2 or O50	UT	1 (0, 1)	−	−	+	−	−	−	−	−	−	−	−	+	−
O5	UT	1 (0, 1)	+	−	−	−	−	−	+	−	−	−	−	+	−
O8	UT	1 (0, 1)	−	−	+	−	−	−	−	−	+	−	−	+	+
O55	O55	1 (1, 0)	+	−	−	−	−	−	−	−	+	−	−	−	−
O84	UT	1 (0, 1)	+	+	−	−	−	−	+	−	−	−	−	+	−
O91	O91	1 (1, 0)	+	−	−	−	−	−	−	−	+	−	−	+	+
O109	UT	1 (1, 0)	−	−	+	−	−	−	+	+	−	−	−	+	−
O109	UT	1 (0, 1)	−	−	+	−	−	−	+	+	−	+	+	+	−
O109	UT	1 (0, 1)	−	−	+	−	−	−	+	−	−	−	−	+	−
O109	UT	1 (0, 1)	−	−	+	−	−	−	−	−	−	−	−	−	−
O109	UT	1 (1, 0)	−	−	+	−	−	+	−	−	−	−	+	+	−
O113	UT	4 (2, 2)	+	+	−	−	+	−	−	−	−	−	+	−	−
O113	UT	2 (0, 2)	−	−	−	−	+	−	−	−	−	−	−	−	−
O113	UT	1 (0, 1)	−	+	−	−	+	−	−	−	−	−	+	−	−
O113	UT	1 (0, 1)	−	−	−	−	+	−	−	+	−	+	+	−	−
O136	O136	1 (0, 1)	+	−	−	−	−	−	−	−	−	−	+	+	−
O136	O136	1 (0, 1)	+	−	−	−	−	−	−	−	−	−	−	+	−
O150	UT	1 (0, 1)	+	−	+	−	−	−	+	−	−	−	−	+	−
O156	UT	1 (1, 0)	+	+	−	−	−	−	+	−	−	−	−	+	−
O157	O157	4 (1, 3)	−	−	+	+	−	−	+	+	−	+	+	+	−
O157	O157	2 (0, 2)	−	−	−	+	−	−	+	+	−	+	+	+	−
O163	UT	1 (0, 1)	+	−	−	−	−	−	−	−	+	−	−	+	+
O168	O168	1 (0, 1)	+	−	−	−	−	−	−	−	−	−	−	−	−
O174	UT	1 (0, 1)	−	−	−	+	−	−	−	−	−	−	−	−	−
O177	UT	1 (1, 0)	−	−	−	+	−	−	+	+	−	−	+	+	−
UT	UT	1 (0, 1)	+	−	+	−	−	−	+	−	−	−	−	+	−

[a] The stx_{1d}, stx_{2b}, stx_{2e}, and stx_{2f} gene variants were not detected. +, positive by PCR; −, negative by PCR.
[b] UT, untypeable.

prevalence rates in cattle of approximately 40 to 70% *(25, 37)*. The differences in these findings may be explained by differences in feed, seasonal peak, age, or detecting methods. Generally, PCR methods disclose a large number of positive samples, but only a portion of those samples yield cultures. In this study, CT-sMAC, DHL, and CHROMagar O26/O157 plates were used to isolate STEC. The CT-sMAC and CHROMagar O26/O157 plates isolated about three-fourths of all the isolates, and DHL plates isolated only two isolates (O2 or O50 and O109; data not shown). Therefore, this study indicates that CT-sMAC and CHROMagar O26/O157 plates are useful in isolating non-O157 STEC. Although 176 broths (64.2%) were *stx* positive by PCR, STEC isolates were cultivated from only 39 samples (22%). Although this result was considerably higher than the 7% previously reported, more than three-fourths of PCR-positive samples were negative by culture-based methods *(7)*.

The STEC O antigen of the isolates was determined by the O serotyping and genotyping methods. Although serotyping offers a very precise and reliable method for differentiating STEC isolates, only a few reference centers in the world are able to provide a full serotyping service. In addition, some field strains of *E. coli* have O antigens that do not react with the respective available antisera or react with more than two antisera. The genotyping method overcomes these problems, and combining both methods identifies the type of O antigen more precisely. In this study, only one isolate (1 of 39) was untyped by using both typing methods. This result was relatively small in number compared with other studies using only the serotyping method *(25, 26, 36)*.

Seventeen types of O antigen were isolated from cattle feces: O1, O2 or O50, O5, O8, O55, O84, O91, O109, O113, O136, O150, O156, O157, O163, O168, O174, and O177. The O113 STEC was the most prevalent O serotype, and other studies have also reported its isolation from cattle samples *(25, 26, 36)*. However, this occurrence has frequently been reported in clinical cases in humans *(10, 27, 34)*. Virulence gene analysis showed that stx_{2d} was common to all O113 STEC isolates, and this result coincided with previous work *(27)*. Although previous studies have shown that O113 STEC isolated from patients with hemolytic uremic syndrome or gastrointestinal illness possessed *hlyA* genes, all the O113 STEC isolates in this study did not contain *hlyA (27, 34)*. Therefore, *hlyA* might be the important virulence factor of STEC O113 against humans. Bacterial adherence and subsequent colonization on the intestinal epithelium cells also contribute to the infection and pathogenicity of STEC. As previously

reported for bovine isolates, more than half of our strains (38.4%, 15 of 39) were *eaeA* (intimin) negative *(25)*. Additionally, *saa* (another adhesion factor) was present in only four isolates. The third most prevalent types of O antigen in this study were O109 and O1. Although some O109 isolates possessed *hlyA* and *eaeA*, this serotype has been mainly isolated from healthy humans and cattle as a carrier. To the best of our knowledge, O109 STEC has not been reported as a cause of human infections. O1 was reported to be the source of some instances of diarrhea in human subjects *(40)*.

We investigated the prevalence of STEC in cattle and identified O serotypes and some virulence factors of isolates. Although many beef cattle possessed STEC, and they are recognized as an important reservoir of STEC, most of the cattle isolates are asymptomatic. This study revealed that many of cattle isolates did not harbor multivirulence factors, such as *hlyA* and some cell adhesion factors. The difference in prevalence of O antigen and virulence-related markers between humans and cattle may promote the identification of the important virulence factors of non-O157 STEC. Furthermore, accumulation of this kind of data may improve the field surveillance or monitoring programs and help find the source of infection or predict the prevalence of STEC in humans.

ACKNOWLEDGMENTS

The authors thank Y. Suehiro and K. Kumamoto (Miyakonojo Meat Inspection Center, Miyazaki Prefecture, Japan) for providing the samples. The work described in this report was funded by the Project for Zoonoses Education and Research.

REFERENCES

1. Armstrong, G. L., J. Hollingsworth, and J. G. Morris. 1996. Emerging foodborne pathogens: *Escherichia coli* O157:H7 as a model of entry of a new pathogen into the food supply of the developed world. *Epidemiol. Rev.* 18:29–51.
2. Barlow, R. S., and G. E. Mellor. 2010. Prevalence of enterohemorrhagic *Escherichia coli* serotypes in Australian beef cattle. *Foodborne Pathog. Dis.* 7:1239–1245.
3. Bettelheim, K. A., A. Kuzevski, R. A. Gilbert, D. O. Krause, and C. S. McSweeney. 2005. The diversity of *Escherichia coli* serotypes and biotypes in cattle faeces. *J. Appl. Microbiol.* 98:699–709.
4. Beutin, L., S. Kaulfuss, T. Cheasty, B. Brandenburg, S. Zimmermann, K. Gleier, G. A. Willshaw, and H. R. Smith. 2002. Characteristics and association with disease of two major subclones of Shiga toxin (Verocytotoxin)-producing strains of *Escherichia coli* (STEC) O157 that are present among isolates from patients in Germany. *Diagn. Microbiol. Infect. Dis.* 44:337–346.
5. Bono, J. L., J. E. Keen, M. L. Clawson, L. M. Durso, M. P. Heaton, and W. W. Laegreid. 2007. Association of *Escherichia coli* O157:H7 *tir* polymorphisms with human infection. *BMC Infect. Dis.* 7:98.
6. 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.
7. Caprioli, A., S. Morabito, H. Brugère, and E. Oswald. 2005. Enterohaemorrhagic *Escherichia coli*: emerging issues on virulence and modes of transmission. *Vet. Res.* 36:289–311.
8. Cobbold, R. N., D. H. Rice, M. Szymanski, D. R. Call, and D. D. Hancock. 2004. Comparison of Shiga-toxigenic *Escherichia coli* prevalences among dairy, feedlot, and cow-calf herds in Washington State. *Appl. Environ. Microbiol.* 70:4375–4378.
9. Currie, A., J. MacDonald, A. Ellis, J. Siushansian, L. Chui, M. Charlebois, M. Peermohamed, D. Everett, M. Fehr and L. K. Ng. 2007. Outbreak of *Escherichia coli* O157:H7 infections associated with consumption of beef donair. *J. Food Prot.* 70:1483–1488.
10. dos Santos, L. F., K. Irino, T. M. Vaz, and B. E. Guth. 2010. Set of virulence genes and genetic relatedness of O113:H21 *Escherichia coli* strains isolated from the animal reservoir and human infections in Brazil. *J. Med. Microbiol.* 59:634–640.
11. Doughty, S., J. Sloan, V. Bennett-Wood, M. Robertson, R. M. Robins-Browne, and E. L. Hartland. 2002. Identification of a novel fimbrial gene cluster related to long polar fimbriae in locus of enterocyte effacement-negative strains of enterohemorrhagic *Escherichia coli*. *Infect. Immun.* 70:6761–6769.
12. Etcheverría, A. I., and N. L. Padola. 2013. Shiga toxin-producing *Escherichia coli*: factors involved in virulence and cattle colonization. *Virulence* 4:366–372.
13. Farfan, M. J., and A. G. Torres. 2012. Molecular mechanisms that mediate colonization of Shiga toxin-producing *Escherichia coli* strains. *Infect. Immun.* 80:903–913.
14. Friedrich, A. W., M. Bielaszewska, W. L. Zhang, M. Pulz, T. Kuczius, A. Ammon, and H. Karch. 2002. *Escherichia coli* harboring Shiga toxin 2 gene variants: frequency and association with clinical symptoms. *J. Infect. Dis.* 185:74–84.
15. Fukushima, H., and M. Gomyoda. 1999. Hydrochloric acid treatment for rapid recovery of Shiga toxin-producing *Escherichia coli* O26, O111 and O157 from faeces, food and environmental samples. *Zentbl. Bakteriol.* 289:285–299.
16. Fukushima, H., K. Hoshina, and M. Gomyoda. 2000. Selective isolation of *eae*-positive strains of Shiga toxin-producing *Escherichia coli*. *J. Clin. Microbiol.* 38:1684–1687.
17. Griffin, P. M., and R. V. Tauxe. 1991. The epidemiology of infections caused by *Escherichia coli* O157:H7, other enterohemorrhagic *E. coli*, and the associated hemolytic uremic syndrome. *Epidemiol. Rev.* 13:60–98.
18. Hilborn, E. D., J. H. Mermin, P. A. Mshar, J. L. Hadler, A. Voetsch, C. Wojtkunski, M. Swartz, R. Mshar, M. A. Lambert-Fair, J. A. Farrar, M. K. Glynn, and L. Slutsker. 1999. A multistate outbreak of *Escherichia coli* O157:H7 infections associated with consumption of mesclun lettuce. *Arch. Intern. Med.* 159:1758–1764.
19. 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.
20. 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.
21. King, L. A., A. Mailles, P. Mariani-Kurkdjian, C. Vernozy-Rozand, M. P. Montet, F. Grimont, N. Pihier, H. Devalk, F. Perret, E. Bingen, E. Espié, and V. Vaillant. 2009. Community-wide outbreak of *Escherichia coli* O157:H7 associated with consumption of frozen beef burgers. *Epidemiol. Infect.* 137:889–896.
22. Kobayashi, H., J. Shimada, M. Nakazawa, T. Morozumi, T. Pohjanvirta, S. Pelkonen, and K. Yamamoto. 2001. Prevalence and characteristics of Shiga toxin-producing *Escherichia coli* from healthy cattle in Japan. *Appl. Environ. Microbiol.* 67:484–489.
23. Locking, M. E., S. J. O'Brien, W. J. Reilly, E. M. Wright, D. M. Campbell, J. E. Coia, L. M. Browning, and C. N. Ramsay. 2001. Risk factors for sporadic cases of *Escherichia coli* O157 infection: the importance of contact with animal excreta. *Epidemiol. Infect.* 127:215–220.
24. Mayer, C. L., C. S. Leibowitz, S. Kurosawa, and D. J. Stearns-Kurosawa. 2012. Shiga toxins and the pathophysiology of hemolytic uremic syndrome in humans and animals. *Toxins* 4:1261–1287.
25. Monaghan, A., B. Byrne, S. Fanning, T. Sweeney, D. McDowell, and D. J. Bolton. 2011. Serotypes and virulence profiles of non-O157 Shiga toxin-producing *Escherichia coli* isolates from bovine farms. *Appl. Environ. Microbiol.* 77:8662–8668.
26. Monaghan, A., B. Byrne, S. Fanning, T. Sweeney, D. McDowell, and D. J. Bolton. 2012. Serotypes and virulotypes of non-O157

shiga-toxin producing *Escherichia coli* (STEC) on bovine hides and carcasses. *Food Microbiol.* 32:223–229.
27. Monaghan, A., B. Byrne, D. McDowell, A. M. Carroll, E. B. McNamara, and D. J. Bolton. 2012. Characterization of farm, food, and clinical Shiga toxin-producing *Escherichia coli* (STEC) O113. *Foodborne Pathog. Dis.* 9:1088–1096.
28. National Institute of Health and Infectious Disease. 2013. Enterohemorrhagic *Escherichia coli* infection in Japan as of April 2013. *Infect. Agents Surveill. Rep.* 34:123–124.
29. Olsen, S. J., G. Miller, T. Breuer, M. Kennedy, C. Higgins, J. Walford, G. McKee, K. Fox, W. Bibb, and P. Mead. 2002. A waterborne outbreak of *Escherichia coli* O157:H7 infections and hemolytic uremic syndrome: implications for rural water systems. *Emerg. Infect. Dis.* 8:370–375.
30. Orth, D., K. Grif, A. B. Khan, A. Naim, M. P. Dierich, and R. Würzner. 2007. The Shiga toxin genotype rather than the amount of Shiga toxin or the cytotoxicity of Shiga toxin in vitro correlates with the appearance of the hemolytic uremic syndrome. *Diagn. Microbiol. Infect. Dis.* 59:235–242.
31. Paton, A. W., and J. C. Paton. 1998. Detection and characterization of Shiga toxigenic *Escherichia coli* by using multiplex PCR assays for stx_1, stx_2, *eaeA*, enterohemorrhagic *E. coli hlyA*, rfb_{O111}, and rfb_{O157}. *J. Clin. Microbiol.* 36:598–602.
32. Paton, A. W., and J. C. Paton. 2002. Direct detection and characterization of Shiga toxigenic *Escherichia coli* by multiplex PCR for stx_1, stx_2, *eae*, *ehxA*, and *saa*. *J. Clin. Microbiol.* 40:271–274.
33. Paton, A. W., P. Srimanote, U. M. Talbot, H. Wang, and J. C. Paton. 2004. A new family of potent AB_5 cytotoxins produced by Shiga toxigenic *Escherichia coli*. *J. Exp. Med.* 200:35–46.
34. Paton, A. W., M. C. Woodrow, R. M. Doyle, J. A. Lanser, and J. C. Paton. 1999. Molecular characterization of a Shiga toxigenic *Escherichia coli* O113:H21 strain lacking *eae* responsible for a cluster of cases of hemolytic-uremic syndrome. *J. Clin. Microbiol.* 37:3357–3361.
35. Paton, J. C., and A. W. Paton. 1998. Pathogenesis and diagnosis of Shiga toxin-producing *Escherichia coli* infections. *Clin. Microbiol. Rev.* 11:450–479.
36. Polifroni, R., A. I. Etcheverría, M. E. Sanz, R. E. Cepeda, A. Krüger, P. M. Lucchesi, D. Fernández, A. E. Parma, and N. L. Padola. 2012. Molecular characterization of Shiga toxin-producing *Escherichia coli* isolated from the environment of a dairy farm. *Curr. Microbiol.* 65:337–343.
37. Pradel, N., V. Livrelli, C. De Champs, J. B. Palcoux, A. Reynaud, F. Scheutz, J. Sirot, B. Joly, and C. Forestier. 2000. Prevalence and characterization of Shiga toxin-producing *Escherichia coli* isolated from cattle, food, and children during a one-year prospective study in France. *J. Clin. Microbiol.* 38:1023–1031.
38. Riley, L. W., R. S. Remis, S. D. Helgerson, H. B. McGee, J. G. Wells, B. R. Davis, R. J. Hebert, E. S. Olcott, L. M. Johnson, N. T. Hargrett, P. A. Blake, and M. L. Cohen. 1983. Hemorrhagic colitis associated with a rare *Escherichia coli* serotype. *N. Engl. J. Med.* 308:681–685.
39. Schmidt, H., L. Beutin, and H. Karch. 1995. Molecular analysis of the plasmid-encoded hemolysin of *Escherichia coli* O157:H7 strain EDL 933. *Infect. Immun.* 63:1055–1061.
40. Srámková, L., M. Bielaszewská, J. Janda, K. Bláhova, and O. Hausner. 1990. Vero cytotoxin-producing strains of *Escherichia coli* in children with haemolytic uraemic syndrome and diarrhoea in Czechoslovakia. *Infection* 18:204–209.
41. Szalo, I. M., F. Goffaux, V. Pirson, D. Piérard, H. Ball, and J. Mainil. 2002. Presence in bovine enteropathogenic (EPEC) and enterohaemorrhagic (EHEC) *Escherichia coli* of genes encoding for putative adhesins of human EHEC strains. *Res. Microbiol.* 153:653–658.
42. Thomas, K. M., M. S. McCann, M. M. Collery, A. Logan, P. Whyte, D. A. McDowell, and G. Duffy. 2012. Tracking verocytotoxigenic *Escherichia coli* O157, O26, O111, O103 and O145 in Irish cattle. *Int. J. Food Microbiol.* 153:288–296.
43. Torres, A. G., J. A. Giron, N. T. Perna, V. Burland, F. R. Blattner, F. Avelino-Flores, and J. B. Kaper. 2002. Identification and characterization of *lpf*ABCC'DE, a fimbrial operon of enterohemorrhagic *Escherichia coli* O157:H7. *Infect. Immun.* 70:5416–5427.
44. Werber, D., A. Fruth, U. Buchholz, R. Prager, M. H. Kramer, A. Ammon, and H. Tschäpe. 2003. Strong association between Shiga toxin-producing *Escherichia coli* O157 and virulence genes stx_2 and *eae* as possible explanation for predominance of serogroup O157 in patients with haemolytic uraemic syndrome. *Eur. J. Clin. Microbiol. Infect. Dis.* 22:726–730.
45. World Health Organization Collaborating Centre for Reference and Research on *Escherichia* and *Klebsiella*. 2012. Identification of three *vtx1* and seven *vtx2* subtypes of Verocytotoxin encoding genes of *Escherichia coli* by conventional PCR amplification Version 6. Available at: http://www.ssi.dk/English/HealthdataandICT/National%20Reference%20Laboratories/~/media/Indhold/EN%20-%20engelsk/Public%20Health/National%20Reference%20Laboratories/vtx%20detection%20%20subtyping%20protocolrev6final.ashx. Accessed 1 December 2012.
46. Yu, J., and J. B. Kaper. 1992. Cloning and characterization of the *eae* gene of enterohaemorrhagic *Escherichia coli* O157:H7. *Mol. Microbiol.* 6:411–417.

J. Gen. Appl. Microbiol., **60**, 44–50 (2014)
doi 10.2323/jgam.60.44
©2014 Applied Microbiology, Molecular and Cellular Biosciences Research Foundation

jGAM

Full Paper

A novel small regulatory RNA enhances cell motility in enterohemorrhagic *Escherichia coli*

(Received November 1, 2013; Accepted November 24, 2013)

Naoki Sudo,[1] Akiko Soma,[2] Akira Muto,[3] Sunao Iyoda,[4] Mayumi Suh,[1] Nanako Kurihara,[1] Hiroyuki Abe,[5,†] Toru Tobe,[5] Yoshitoshi Ogura,[6] Tetsuya Hayashi,[6] Ken Kurokawa,[7] Makoto Ohnishi,[4] and Yasuhiko Sekine[1,*]

[1] *Department of Life Science, College of Science, Rikkyo University, Toshima, Tokyo 171–8501, Japan*
[2] *Department of Horticulture, Chiba University, Matsudo, Chiba 271–8510, Japan*
[3] *Department of Biochemistry and Molecular Biology, Faculty of Agriculture and Life Science, Hirosaki University, Hirosaki, Aomori 036–8561, Japan*
[4] *Department of Bacteriology, National Institute of Infectious Diseases, Shinjuku, Tokyo 162–8640, Japan*
[5] *Department of Biomedical Informatics, Osaka University Graduate School of Medicine, Suita, Osaka 565–0871, Japan*
[6] *Division of Microbial Genomics, Department of Genomics and Bioenvironmental Science, Frontier Science Research Center, University of Miyazaki, Kiyotake, Miyazaki 889–1692, Japan*
[7] *Earth-Life Science Institute, Tokyo Institute of Technology, Yokohama, Kanagawa 226–8501, Japan*

Small regulatory RNAs (sRNAs) are conserved among a wide range of bacteria. They modulate the translational efficiency of target mRNAs through base-pairing with the help of RNA chaperone Hfq. The present study identified a novel sRNA, Esr41 (enterohemorrhagic *Escherichia coli* O157 small RNA #41), from an intergenic region of an enterohemorrhagic *E. coli* (EHEC) O157:H7 Sakai-specific sequence that is not present in the nonpathogenic *E. coli* K-12. Esr41 was detected as an RNA molecule approximately 70 nucleotides long with a 3′ GC-rich palindrome sequence followed by a long poly(U), which is a characteristic of rho-independent terminators and is also a structural feature required for the action of Hfq. EHEC O157 harboring a multicopy plasmid carrying the *esr41* gene increased cell motility and the expression of *fliC*, a gene encoding a major flagellar component. These results indicate that Esr41 stimulates *fliC* expression in EHEC O157. Furthermore, the increase in cell motility induced by Esr41 was also observed in the *E. coli* K-12, suggesting that target genes controlled by Esr41 are present in both EHEC O157 and K-12.

Key words : cell motility; EHEC O157 : H7 Sakai; flagella; small regulatory RNA

Introduction

Small regulatory RNAs (sRNAs) play crucial roles in bacterial stress responses and virulence (Gottesman, 2004; Papenfort and Vogel, 2010; Waters and Storz, 2009). Many sRNAs are induced under specific physiological conditions and control the expression of target genes, primarily at the post-transcriptional levels (Gottesman, 2004; Wassarman, 2002). sRNAs are classified as *cis*-acting sRNAs or *trans*-acting sRNAs. *Cis*-acting sRNAs are encoded on the strand opposite to the sense strand of the target gene and regulate the target gene expression by perfect base-pairing with its mRNA (Kawano et al., 2005). *Trans*-acting sRNAs are usually encoded in a different region from that of the target genes and regulate the expression of the target genes by partial base-pairing with their mRNAs. One major class of sRNA binds to RNA chaperone Hfq, which facilitates base-pairing between sRNA and the target mRNA to regulate its translation, mostly negatively but also positively in some cases (Aiba, 2007; Vogel and Luisi, 2011).

A comprehensive identification of sRNA has been performed in many bacteria, such as *Escherichia coli* and

*Corresponding author: Yasuhiko Sekine, Department of Life Science, College of Science, Rikkyo University, Toshima, Tokyo 171–8501, Japan
Email: ysekine@rikkyo.ac.jp
†Present address: Department of Molecular Bacteriology, Research Institute for Microbial Diseases, Osaka University, Suita, Osaka 565–0871, Japan

None of the authors of this manuscript has any financial or personal relationship with other people or organizations that could inappropriately influence their work.

Salmonella spp. (Huang et al., 2009; Sharma and Vogel, 2009). Expansive and diverse searches for sRNA have been carried out using several experimental methods: shotgun cloning (Kawano et al., 2005; Vogel et al., 2003), co-immunoprecipitation with Hfq (Sittka et al., 2008; Zhang et al., 2003), and deep sequencing (Liu et al., 2009; Shinhara et al., 2011; Sittka et al., 2008). These analyses have identified at least 110 sRNAs in *E. coli* K-12, which have been registered in RegulonDB 8.3 (Salgado et al., 2013). Some of these have also been well characterized (Majdalani et al., 1998; Masse and Gottesman, 2002; Mizuno et al., 1984; Møller et al., 2002; Vanderpool and Gottesman, 2004). However, most of these searches were performed in the nonpathogenic *E. coli* K-12, and almost no investigations have been undertaken in pathogenic *E. coli*, including enterohemorrhagic *E. coli* (EHEC). EHEC is an important class of diarrheagenic *E. coli* associated with severe diseases, such as hemorrhagic colitis and hemolytic uremic syndrome (Frankel et al., 1998; Nataro and Kaper, 1998). Though the most important virulence determinant responsible for severe illness is Shiga toxin, other virulence factors, including the Type 3 (T3) secretion system and T3 effectors, also contribute to the EHEC pathogenicity (Elliott et al., 1998; Frankel et al., 1998; Nataro and Kaper, 1998). The expression of these virulence factors is strictly regulated through a complicated and only partially understood regulatory network (Elliott et al., 2000; Honda et al., 2009; Iyoda and Watanabe, 2004; Iyoda et al., 2006; Sperandio et al., 2003). Despite the detailed knowledge of sRNAs/Hfq-mediated regulation in the *E. coli* K-12, little is known about the contribution of sRNA to regulation in EHEC. On the other hand, it was reported that EHEC *hfq* deletion mutants massively overproduce and secrete T3 effectors encoded on the locus of enterocyte effacement (LEE) or elsewhere in the chromosome, indicating that Hfq negatively regulates the expression of many virulence-associated genes (Hansen and Kaper, 2009; Shakhnovich et al., 2009). These observations led us to suppose that sRNA/Hfq-mediated regulation is important for pathogenicity in EHEC.

An EHEC strain, O157:H7 Sakai (referred to as O157 Sakai), was isolated from a typical patient during an outbreak in Sakai City, Osaka, Japan in 1996. Its chromosome is approximately 5.5 Mb in size: 4.1 Mb of the sequence is conserved between O157 Sakai and the *E. coli* K-12. The remaining 1.4 Mb of the sequence is present in O157 Sakai but not in K-12, and contains many prophages and integrative elements (Hayashi et al., 2001b). Most virulence-associated genes are encoded within the O157 Sakai-specific sequence (Tobe et al., 2006). Although the O157 Sakai genome encodes 95 sRNAs that are also present in K-12, sRNAs encoded in the O157 Sakai-specific sequence are not well defined.

In this study, we identified a novel sRNA from an intergenic region within the O157 Sakai-specific sequence, and showed that this sRNA enhances cell motility in both O157 Sakai and *E. coli* K-12, suggesting that it controls the expression of motility-related genes that are present in both O157 Sakai and K-12.

Materials and Methods

Bacterial strains, plasmids, and media. The *E. coli* strains and oligonucleotides used in this study are listed in Table 1 and Table 2, respectively. SKI-5142 is a *lac*-negative derivative of the EHEC O157:H7 Sakai (Iyoda and Watanabe, 2005) and was used as the wild-type strain. SKI-5350, which contains a chromosomal *fliC'-lacZ* translational fusion gene, was constructed as described previously (Saitoh et al., 2008) using the primer pair sets, FLIC-LACZf1-TL / LACZ-FRTr1 and LACZ-FRTf1 / FLIC-FRTr1.

SKI-5142Δ*esr41* was constructed from SKI-5142 by a one-step gene inactivation protocol (Datsenko and Wanner, 2000), using the PCR primer pair, #41-pKD13-5′ and #41-pKD13-3′. MG1655Δ*lac* was constructed by P1 transduction of MG1655 with P1 phage grown on W3110Δ*lac* and selecting for chloramphenicol resisitance.

Plasmid pGEM-SR41-F carries the *esr41* coding region and its putative promoter sequence. It was constructed by PCR amplification of a SKI-5142 fragment using the primers, #41temp5′-1 and #41temp3′, followed by cloning the amplified fragment into pGEM-T-easy vector (Promega). Plasmid pRS-Esr41 was constructed by inserting an *Eco*RI fragment of pGEM-SR41-F, which contains the *esr41* coding region and its putative promoter sequence, into an *Eco*RI site of pRS414 (Simons et al., 1987).

E. coli cells were grown aerobically at 37°C in LB medium or Dulbecco's modified Eagle's medium (DMEM, Invitrogen) supplemented with ampicillin (100 μg ml^{-1}), chloramphenicol (25 μg ml^{-1}), or kanamycin (50 μg ml^{-1}). DMEM and TB (10 g L^{-1} of Tryptone and 0.5 g L^{-1} of NaCl) motility agar plates were prepared by adding 0.25% agar (Shoei) to DMEM.

RNA isolation and Northern blotting. Total RNA was prepared from *E. coli* cells grown at 37°C in LB by the acid guanidinium thiocyanate-phenol-chloroform method using ISOGEN (Nippon Gene) according to the manufacturer's

Table 1. *E. coli* strains used in this study.

Name	Relevant genotype and property	Source
EHEC derivatives		
SakaiΔ*stx*	SakaiΔ*stx1,2*	Gift from Sasakawa laboratory
SKI-5142	SakaiΔ(*lacIZYA*)	Iyoda and Watanabe, 2005
SKI-5142Δ*esr41*	SKI-5142 Δ*esr41*::Km^r	This study
SKI-5350	SKI-5142 *fliC'-lacZ* translational fusion, Cm^r	This study
E. coli K-12 derivatives		
W3110		Laboratory stock
MG1655	Same as CGSC#6300	CGSC
W3110Δ*lac*	W3110 Δ*lacZ*::Cm^r	Gift from K. Tanaka
MG1655Δ*lac*	MG1655 Δ*lacZ*::Cm^r	This study

Table 2. Oligonucleotides used in this sutudy.

Name	Sequence (5′→3′)
FLIC-LACZf1-TL	GCAATATAGGATAACGAATCATGGCACAAGTCATTAATACCAACAGCCTCGTCGTTTTACAACGTCGTGA
FLIC-FRTr1	ATCAGGCAATTTGGCGTTGCCGTCAGTCTCAGTTAATCAGGTTACAACGACATATGAATATCCTCCTTAG
LACZ-FRTf1	TGGTCTGGTGTCAAAAATAAGTGTAGGCTGGAGCTGCTTC
LACZ-FRTr1	GAAGCAGCTCCAGCCTACACTTATTTTTGACACCAGACCA
#41-pKD13-5′	CCCGGAGCAAATTCAGGTAGGTCATGCATAAAAATATCTTTCTATACGTGTAGGCTGGAGCTGCTTC
#41-pKD13-3′	CGAGGTATGTTTCTGGTGGAAAGGACAGTGGGATTAAAAAGTAAGATTCCGGGGATCCGTCGACC
#41temp5′-1	GCAGAGATGTTGATGGCGGAA
#41temp3′	CTGGTGGAAAGGACAGTGGG
Oligo-41	TAGTGCCGGCGGAGGGAATACCCCATGGAG

protocol. The RNA was then separated on a 6% polyacrylamide gel containing 7 M urea and compared to RNA molecular weight markers (RNA Size Standard Marker III, Wako). The gel was blotted onto a Hybond N$^+$ membrane (GE Healthcare), and incubated with the ^{32}P-labeled oligonucleotide Oligo-41 at 50°C for 6 h. The membrane was then extensively washed with 1 × SSC buffer (150 mM NaCl and 15 mM sodium citrate) at room temperature for 30 min, exposed to a phosphorimaging plate, and scanned using the Molecular Dynamics Typhoon (Model 9210).

Primer extension. Total RNA was prepared as described above. After heating at 95°C for 2 min, the total RNA (6 μg) was allowed to anneal, on ice, to the primer, ^{32}P-labeled Oligo-41 (0.2 pmol). Reverse transcription reactions were carried out using ReverTraAce (TOYOBO) at 55°C for 30 min and stopped by heating at 95°C for 2 min and adding the stop solution (TOYOBO). A sequencing ladder was generated using PCR product amplified from pRS-Esr41 using the primers #41temp5′-1 and #41temp3′, ^{32}P-labeled Oligo-41, and Sequencing Pro (TOYOBO), according to the manufacturer's instructions. Samples were denatured at 95°C for 3 min prior to separation on an 8% polyacrylamide gel containing 8 M urea. The gel was then dried and exposed to a phosphorimaging plate, and scanned using the Molecular Dynamics Typhoon (Model 9210).

Motility assay. To measure bacterial motility, *E. coli* cells grown on LB agar plates were spotted onto DMEM or TB motility agar plates and incubated at 37°C for 16–48 h. All assays were repeated at least three times.

Western blotting. *E. coli* cells were grown in LB or DMEM at 37°C with shaking until they reached an optical density of 0.8 at 600 nm. Proteins in whole-cell lysates prepared from 0.75 ml of the cell culture were analyzed by Western blotting as described previously (Iyoda and Kutsukake, 1995; Iyoda and Watanabe, 2004). Western blotting was performed using the polyclonal antibody anti-H7 (FliC) (Denka Seiken), and the bound antibody was detected using ECL Western Blotting detection reagents (GE Healthcare) as described previously (Iyoda and Watanabe, 2004). All assays were performed in duplicate and were repeated at least three times.

Assay for β-galactosidase activity. β-Galactosidase activity was assayed as described previously (Iyoda and Kutsukake, 1995; Iyoda and Watanabe, 2004). *E. coli* cells grown in LB or DMEM at 37°C with shaking were harvested at an optical density of 0.8 at 600 nm. All assays were performed in triplicate and were repeated at least three times.

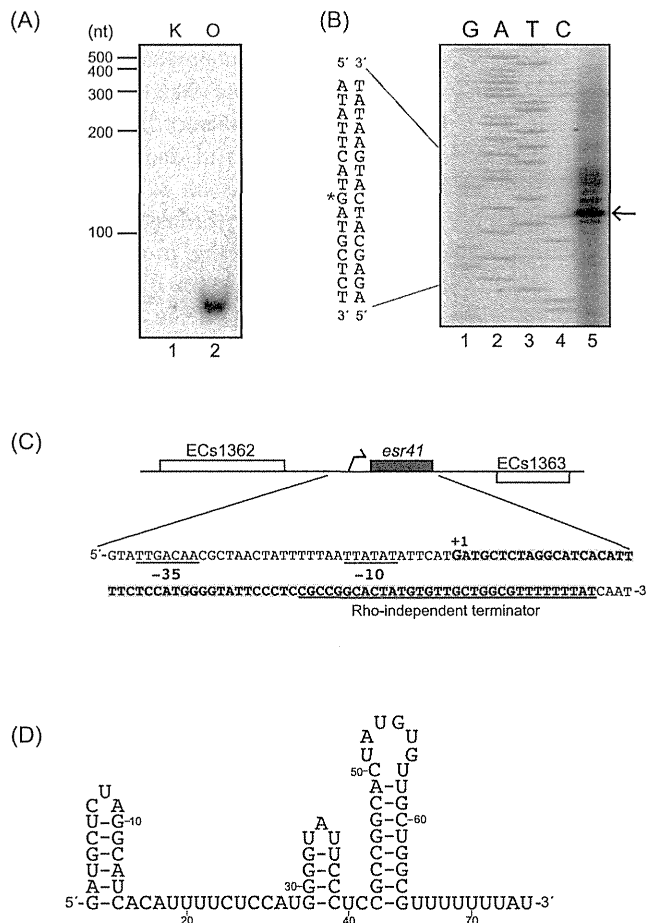

Fig. 1. Identification of Esr41.

(A) Northern blot analysis to identify Esr41. Total RNAs prepared from *E. coli* K-12 W3110 strain (lane 1; labeled K) and O157 Sakai (lane 2; labeled O) were subjected to Northern blot analysis using Oligo-41. Positions of the RNA size markers are indicated on the left. (B) Mapping of the 5′ end of Esr41 by primer extension. Primer extension analysis using the ^{32}P-labeled primer Oligo-41 was performed using total RNA from O157 Sakai grown in LB. A sequence ladder (G, A, T, and C) was prepared using the same primer, but with the pRS-Esr41 as the plasmid template, and run in parallel. The arrow indicates the elongated primer band, which corresponds to the 5′ end of Esr41. An asterisk indicates the mapped 5′ end of Esr41. (C) Genomic context of *esr41* gene. The gene encoding Esr41 is located between ECs1362 and ECs1363. The coding sequence of Esr41 is shaded grey. The putative −10 and −35 promoter elements and rho-independent terminator are indicated by underlines. (D) Nucleotide sequence (accession number AB861978) and possible secondary structure of Esr41 as predicted by mfold (Zuker, 2003).

Results and Discussion

Identification of a novel small RNA from EHEC O157:H7 Sakai genome

The O157 Sakai-specific sequences contain a total of

1,632 protein-coding sequences (CDSs) (Hayashi et al., 2001b) and 115 intergenic regions (IGRs) longer than 400 bp. We numbered these from 1 to 115. By searching the IGRs for sRNA-encoding genes based on the presence of promoter- and/or terminator-like sequences, 27 candidate sequences were identified. Northern blot analysis of the total RNA isolated from O157 Sakai using an oligonucleotide probe hybridizing to IGR 41, which was located between two CDSs (ECs1362 and ECs1363), was performed. As a result, we detected an RNA molecule shorter than 100 nucleotides [nt] (Fig. 1A), which was thereafter called Esr41 (enterohemorrhagic *E. coli* O157 small RNA #41). A primer extension experiment identified a major 5′ end of Esr41 (Fig. 1B), and a putative sigma-70 type promoter element was found in a region upstream of the identified 5′ end (Fig. 1C). The 3′ region of the putative coding region of Esr41 contained a possible rho-independent type terminator sequence (Fig. 1C). Judging from the position of the 5′ end and the possible terminator sequence, Esr41 was expected to be 66–74 nt long, which was consistent with the size of RNA detected in the Northern blot analysis (Fig. 1A). The possible rho-independent type terminator contains seven consecutive U residues and subsequent A and U residues (Fig. 1D). Most Hfq-binding sRNAs share this feature; that is, a sequence likely to form a stem-loop followed by a long poly(U) sequence (Otaka et al., 2011; Sauer and Weichenrieder, 2011). This long poly(U) sequence of sRNAs is essential for functional Hfq-binding (Otaka et al., 2011). It is therefore possible that Esr41 is involved in gene expression control as a *trans*-acting Hfq-binding sRNA.

The *esr41* coding region is located within Sakai prophage-like element 1 (SpLE1), a large integrated element. Sequences identical to the O157 Sakai sequence encoding the *esr41* gene and its putative promoter are present not only in the genome of EHECs of serotype O157:H7 (accession number NC_002655.2), O26: H11 (accession number NC_013361), and O111:H− (accession number NC_013364), but also in the genomes of *Shigella flexneri* (accession number NC_008258) and *S. boydii* (accession number NC_010658) (data not shown). Here we should mention that the specific distribution of *esr41* in O157:H7, O26: H11, and O111:H−among sequenced *E. coli* strains is consistent with our previous finding that SpLE1 or SpLE1-like elements are specifically distributed to EHEC O157:H7, O26: H11, and O111:H− (Kusumoto et al., 2011).

A multicopy plasmid carrying the esr41 gene increased cell motility in EHEC

To investigate the function of Esr41, the *esr41* coding region and its putative promoter sequence were cloned into a high copy-number vector plasmid, pRS414; this plasmid was designated pRS-Esr41. EHEC SKI-5142 harboring either pRS-Esr41 or pRS414 (control) was then grown on motility agar plates composed of DMEM, where flagellar synthesis is repressed (Girón et al., 2002; Zhou et al., 2003). EHEC SKI-5142 harboring pRS-Esr41 exhibited significantly increased motility compared with that harboring the pRS414 control vector (Fig. 2A). To examine the effect of deletion of the *esr41* gene on cell motility, we constructed a derivative of SKI-5142 in which the coding region of *esr41* was completely deleted. However, deletion of *esr41* did not affect the cell motility of EHEC SKI-5142 (Fig. 2A).

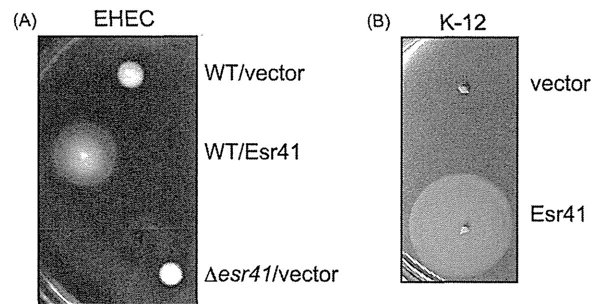

Fig. 2. Effect of multicopy *esr41* gene and *esr41* deletion on motility of O157 Sakai and K-12.

(A) EHEC SKI-5142 harboring pRS414 (WT/vector) or pRS-Esr41 (WT/Esr41) and SKI-5142Δ*esr41* harboring pRS414 (Δ*esr41*/vector) were grown overnight at 37°C on DMEM motility agar plates. (B) *E. coli* K-12 MG1655Δ*lac* strains harboring pRS414 (vector) or pRS-Esr41 (Esr41) were grown for 40 h at 37°C on DMEM motility agar plates.

A multicopy plasmid carrying the esr41 gene increased the expression of fliC in EHEC

Since Esr41 promoted cell motility in DMEM as described above, we next examined the effect of Esr41 on flagellar synthesis. For this purpose, the cellular content of FliC protein, a major component of the flagellar filament, was measured by Western blot analysis. The amount of FliC protein in the SKI-5142 harboring pRS-Esr41 grown in DMEM was significantly increased, compared with the strain harboring the pRS414 control vector (Fig. 3A). A similar result was obtained from cells grown in LB, although the degree of increase was somewhat lower (Fig. 3A). The *esr41* deletion had no effect on the amount of FliC protein when grown in either DMEM or LB (data not shown).

To examine whether the increased amount of FliC was caused by enhanced expression of *fliC*, we first needed to monitor the level of *fliC* expression. For this purpose, we constructed a derivative of the SKI-5142 strain in which the chromosomal *fliC* coding region from the 11th codon to the last codon was replaced with a promoter-less *lacZ* gene. Cells of the resulting strain (SKI-5350), harboring pRS-Esr41, with translational fusion of *fliC* with *lacZ*, expressed β-galactosidase activity that was about 13-fold higher than that in the same strain harboring pRS414, when the cells were grown in DMEM (Fig. 3B). This result indicates that Esr41 enhances *fliC* expression. It is known that the expression level of *fliC* in EHEC O157 grown in LB is higher than that grown in DMEM. Consistent with this, β-galactosidase activity from the *fliC-lacZ* gene in the cells harboring pRS414 control vector grown in LB was about 50-fold higher than that in the cells grown in DMEM. Even in such a higher background *fliC* expression, however, about 2.6-fold enhancement of the *fliC* expression by Esr41 was observed (Fig. 3B). As the enhancing effects of Esr41 on the *fliC* expression were observed at different basal levels of its expression, it is likely that Esr41 does not merely derepress the flagellar regulon in DMEM.

With regard to the increased *fliC* expression in LB, the cells grown on motility agar plates composed of TB, which is a medium similar to LB, showed significantly increased motility compared with the cells grown on DMEM motility

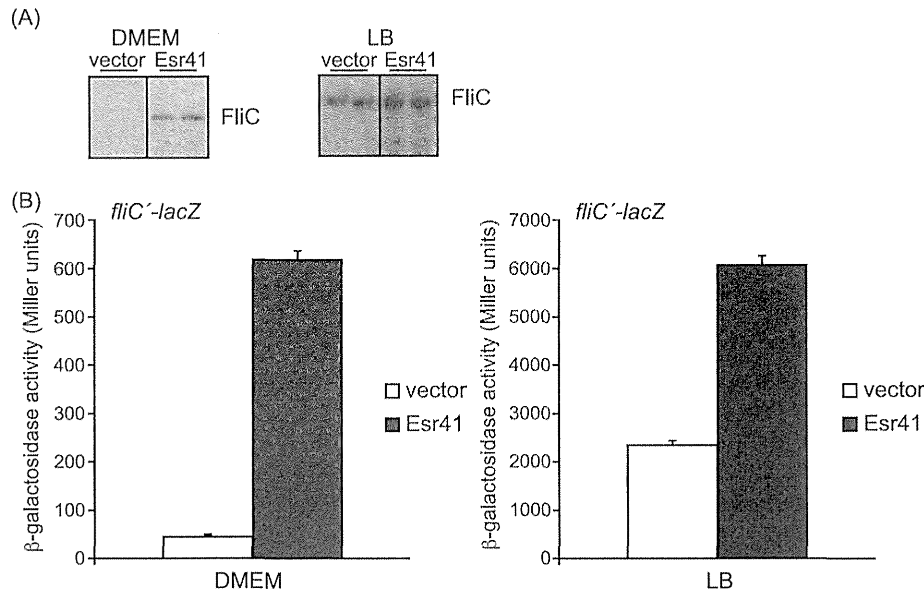

Fig. 3. Effect of multicopy *esr41* gene on the expression of *fliC* and *fliC′-lacZ* in O157 Sakai.

(A) FliC in whole-cell lysates of EHEC SKI-5142 harboring pRS414 (vector) or pRS-Esr41 (Esr41) grown in DMEM or LB were quantified by Western blot analysis using polyclonal anti-FliC (anti-H7) antibody. Samples were prepared from two bacterial cell cultures and separately loaded on the gel side by side. (B) β-Galactosidase activities from the *fliC′-lacZ* translational fusion gene in cultures of the SKI-5350 harboring pRS414 (vector) or that harboring pRS-Esr41 (Esr41), grown in DMEM and LB.

agar plates, but further increase in motility by Esr41 was not observed (data not shown).

A multicopy plasmid carrying the esr41 gene increased cell motility in the E. coli K-12

The genes encoding flagellar components and their regulation factors comprise a large and complex regulon. The flagellar regulon includes three hierarchical classes, Classes 1, 2, and 3, in *E. coli* and *Salmonella enterica* (Chilcott and Hughes, 2000; Kutsukake et al., 1990; Liu and Matsumura, 1994). It is highly conserved in both EHEC and the *E. coli* K-12. This fact led us to hypothesize that Esr41 would enhance cell motility in the *E. coli* K-12 as well. To test this hypothesis, the effect of Esr41 on cell motility in the *E. coli* K-12 was examined on DMEM motility agar plates. As expected, the K-12 strain harboring pRS-Esr41 showed significantly higher cell motility compared with that harboring pRS414, just as the O157 Sakai had (Fig. 2B). This suggests that Esr41 may control the expression of one or more genes that are involved in the regulation of flagellar synthesis in both the O157 Sakai and K-12.

Our results indicate the possibility that Esr41 acts as an important activator in the motility regulation network. Since Esr41 has a feature of *trans*-acting Hfq-binding sRNA, it may regulate the expression of motility-related genes through base-pairing, with the help of Hfq. Eight kinds of sRNAs participate in motility regulation in *E. coli* K-12, and five of these regulate the translation of *flhDC*, which encodes a central activator of the flagellar regulon, through direct base-pairing with the 5′ UTR of *flhDC* mRNA (De Lay and Gottesman, 2012; Thomason et al., 2012). *flhDC* expression is also regulated by various transcription factors (Prüss et al., 2006). In this way, the expression of motility-related genes, including *flhDC*, is strictly regulated at the transcriptional and post-transcriptional levels in response to various environmental cues (Prüss et al., 2006). Our preliminary results showed that overexpression of the *esr41* gene increased *flhD* expression, and a possible base-pairing region between Esr41 and the 5′ UTR of the *flhDC* mRNA is present. Deletion of the possible base-pairing region, however, did not abolish the effect of Esr41 on *flhD* expression (data not shown). Thus our results suggest that Esr41 might interact with the *flhDC* mRNA in a region other than the 5′ UTR or an mRNA encoding a transcription factor that regulates *flhDC* expression.

Cell motility provides some advantages to pathogenic bacteria because it helps pathogenic bacteria reach preferred niches or target tissues/cells and also help them escape from non-preferred environments. On the other hand, bacterial cells expressing extracellular flagellar filaments have a disadvantage because flagellin, the *fliC*-encoded major subunit of flagella, is highly antigenic and activates toll-like receptor 5, leading to a pro-inflammatory response in the host (Hayashi et al., 2001a). It is also known that constitutive expression of *flhDC* greatly decreases efficient adhesion in EHEC (Iyoda et al., 2006). Thus, flagellar expression must be strictly and properly regulated in each phase of infection. Our finding raises the possibility that control of flagellar expression by Esr41 is also important for the establishment of EHEC infection.

Acknowledgments

This work was supported by the Frontier Project Adaptation and Evolution of Extremophiles (to Y.S.), by the Strategic Research Foundation Grant-aided Project for Private Universities (S1201003 to Y.S.) from the Ministry of Education, Culture, Sports, Science, and Technology, Japan, and by the Institute for Fermentation, Osaka (to A.S.).

References

Aiba, H. (2007) Mechanism of RNA silencing by Hfq-binding small RNAs. *Curr. Opin. Microbiol.*, **10**, 134–139.

Chilcott, G. S. and Hughes, K. T. (2000) Coupling of flagellar gene expression to flagellar assembly in *Salmonella enterica* serovar *typhimurium* and *Escherichia coli*. *Microbiol. Mol. Biol. Rev.*, **64**, 694–708.

Datsenko, K. A. and Wanner, B. L. (2000) One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc, Natl, Acad, Sci, USA*, **97**, 6640–6645.

De Lay, N. and Gottesman, S. (2012) A complex network of small non-coding RNAs regulate motility in *Escherichia coli*. *Mol. Microbiol.*, **86**, 524–538.

Elliott, S. J., Sperandio, V., Girón, J. A., Shin, S., Mellies, J. L. et al. (2000) The locus of enterocyte effacement (LEE)-encoded regulator (Ler) controls expression of both LEE- and non-LEE-encoded virulence factors in enteropathogenic and enterohemorrhagic *Escherichia coli*. *Infect. Immun.*, **68**, 6115–6126.

Elliott, S. J., Wainwright, L. A., McDaniel, T. K., Jarvis, K. G., Deng, Y. K. et al. (1998) The complete sequence of the locus of enterocyte effacement (LEE) from enteropathogenic *Escherichia coli* E2348/69. *Mol. Microbiol.*, **28**, 1–4.

Frankel, G., Phillips, A. D., Rosenshine, I., Dougan, G., Kaper, J. B. et al. (1998) Enteropathogenic and enterohaemorrhagic *Escherichia coli*: More subversive elements. *Mol. Microbiol.*, **30**, 911–921.

Girón, J. A., Torres, A. G., Freer, E., and Kaper, J. B. (2002) The flagella of enteropathogenic *Escherichia coli* mediate adherence to epithelial cells. *Mol. Microbiol.*, **44**, 361–379.

Gottesman, S. (2004) The small RNA regulators of *Escherichia coli:* roles and mechanisms. *Annu. Rev. Microbiol.*, **58**, 303–328.

Hansen, A. M. and Kaper, J. B. (2009) Hfq affects the expression of the LEE pathogenicity island in enterohaemorrhagic *Escherichia coli*. *Mol. Microbiol.*, **73**, 446–465.

Hayashi, F., Smith, K. D., Ozinsky, A., Hawn, T. R., Eugene, C. Y. et al. (2001a) The innate immune response to bacterial flagellin is mediated by Toll-like receptor 5. *Nature*, **410**, 1099–1103.

Hayashi, T., Makino, K., Ohnishi, M., Kurokawa, K., Ishiki, K. et al. (2001b) Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genomic comparison with a laboratory strain K-12. *DNA Res.*, **8**, 11–22.

Honda, N., Iyoda, S., Yamamoto, S., Terajima, J., and Watanabe, H. (2009) LrhA positively controls the expression of the locus of enterocyte effacement genes in enterohemorrhagic *Escherichia coli* by differential regulation of their master regulators PchA and PchB. *Mol. Microbiol.*, **74**, 1393–1411.

Huang, H. Y., Chang, H. Y., Chou, C. H., Tseng, C. P., Ho, S. Y. et al. (2009) sRNAMap: Genomic maps for small non-coding RNAs, their regulators and their targets in microbial genomes. *Nucleic Acids Res.*, **37** (Database), D150–154.

Iyoda, S. and Kutsukake, K. (1995) Molecular dissection of the flagellum-specific anti-sigma factor, FlgM, of *Salmonella typhimurium*. *Mol. Gen. Genet.*, **249**, 417–424.

Iyoda, S., Koizumi, N., Satou, H., Lu, Y., Saitoh, T. et al. (2006) The GrlR-GrlA regulatory system coordinately controls the expression of flagellar and LEE-encoded type III protein secretion systems in enterohemorrhagic *Escherichia coli*. *J. Bacteriol.*, **188**, 5682–5692.

Iyoda, S. and Watanabe, H. (2004) Positive effects of multiple *pch* genes on expression of the locus of enterocyte effacement genes and adherence of enterohaemorrhagic *Escherichia coli* O157:H7 to HEp-2 cells. *Microbiology*, **150**, 2357–2571.

Iyoda, S. and Watanabe, H. (2005) ClpXP protease controls expression of the type III protein secretion system through regulation of RpoS and GrlR levels in enterohemorrhagic *Escherichia coli*. *J. Bacteriol.*, **187**, 4086–4094.

Kawano, M., Reynolds, A. A., Miranda-Rios, J., and Storz, G. (2005) Detection of 5′- and 3′-UTR-derived small RNAs and *cis*-encoded antisense RNAs in *Escherichia coli*. *Nucleic Acids Res.*, **33**, 1040–1050.

Kusumoto, M., Ooka, T., Nishiya, Y., Ogura, Y., Saito, T. et al. (2011) Insertion sequence-excision enhancer removes transposable elements from bacterial genomes and induces various genomic deletions. *Nat. Commun.*, **2**, 152.

Kutsukake, K., Ohya, Y., and Iino, T. (1990) Transcriptional analysis of the flagellar regulon of *Salmonella typhimurium*. *J. Bacteriol.*, **172**, 741–747.

Liu, J. M., Livny, J., Lawrence, M. S., Kimball, M. D., Waldor, M. K. et al. (2009) Experimental discovery of sRNAs in *Vibrio cholerae* by direct cloning, 5S/tRNA depletion and parallel sequencing. *Nucleic Acids Res.*, **37**, e46.

Liu, X. and Matsumura, P. (1994) The FlhD/FlhC complex, a transcriptional activator of the *Escherichia coli* flagellar class II operons. *J. Bacteriol.*, **176**, 7345–7351.

Majdalani, N., Cunning, C., Sledjeski, D., Elliott, T., and Gottesman, S. (1998) DsrA RNA regulates translation of RpoS message by an anti-antisense mechanism, independent of its action as an antisilencer of transcription. *Proc. Natl. Acad. Sci. USA*, **95**, 12462–12467.

Masse, E. and Gottesman, S. (2002) A small RNA regulates the expression of genes involved in iron metabolism in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA*, **99**, 4620–4625.

Mizuno, T., Chou, M. Y., and Inouye, M. (1984) A unique mechanism regulating gene expression: Translational inhibition of a complementary RNA transcript (micRNA). *Proc. Natl. Acad. Sci. USA*, **81**, 1966–1970.

Møller, T., Franch, T., Udesen, C., Gerdes, K., and Valentin-Hansen, P. (2002) Spot42 RNA mediates discoordinate expression of the *E. coli* galactose operon. *Genes Dev.*, **16**, 1696–1706.

Nataro, J. P. and Kaper, J. B. (1998) Diarrheagenic *Escherichia coli*. *Clin. Microbiol. Rev.*, **11**, 142–201.

Otaka, H., Ishikawa, H., Morita, T., and Aiba, H. (2011) PolyU tail of rho-independent terminator of bacterial small RNAs is essential for Hfq action. *Proc. Natl. Acad. Sci. USA*, **108**, 13059–13064.

Papenfort, K. and Vogel, J. (2010) Regulatory RNA in bacterial pathogens. *Cell Host. Microbe.*, **8**, 116–127.

Prüss, B. M., Besemann, C., Denton, A., and Wolfe, A. J. (2006) A complex transcription network controls the early stages of biofilm development by *Escherichia coli*. *J. Bacteriol.*, **188**, 3731–3739.

Sauer, E. and Weichenrieder, O. (2011) Structural basis for RNA 3′-end recognition by Hfq. *Proc. Natl. Acad. Sci. USA*, **108**, 13065–13070.

Saitoh, T., Iyoda, S., Yamamoto, S., Lu, Y., Shimuta, K. et al. (2008) Transcription of the *ehx* enterohemolysin gene is positively regulated by GrlA, a global regulator encoded within the locus of enterocyte effacement in enterohemorrhagic *Escherichia coli*. *J. Bacteriol.*, **190**, 4822–4830.

Salgado, H., Peralta-Gil, M., Gama-Castro, S., Santos-Zavaleta, A., Muñiz-Rascado, L. et al. (2013) RegulonDB v8.0: Omics data sets, evolutionary conservation, regulatory phrases, cross-validated gold standards and more. *Nucleic Acids Res.*, **41** (Database issue), D203–13.

Shakhnovich, E. A., Davis, B. M., and Waldor, M. K. (2009) Hfq negatively regulates type III secretion in EHEC and several other pathogens. *Mol. Microbiol.*, **74**, 347–363.

Sharma, C. M. and Vogel, J. (2009) Experimental approaches for the discovery and characterization of regulatory small RNA. *Curr. Opin. Microbiol.*, **12**, 536–546.

Shinhara, A., Matsui, M., Hiraoka, K., Nomura, W., Hirano, R. et al. (2011) Deep sequencing reveals as-yet-undiscovered small RNAs in *Escherichia coli*. *BMC Genom.*, **12**, 428.

Simons, R. W., Houman, F., and Kleckner, N. (1987) Improved single and multicopy lac-based cloning vectors for protein and operon fusions. *Gene*, **53**, 85–96.

Sittka, A., Lucchini, S., Papenfort, K., Sharma, C. M., Rolle, K. et al. (2008) Deep sequencing analysis of small noncoding RNA and mRNA targets of the global post transcriptional regulator, Hfq. *PLoS Genet.*, **4**, e1000163.

Sperandio, V., Torres, A. G., Jarvis, B., Nataro, J. P., and Kaper, J. B. (2003) Bacteria-host communication: The language of hormones. *Proc. Natl. Acad. Sci. USA*, **100**, 8951–8956.

Thomason, M. K., Fontaine, F., De Lay, N., and Storz, G. (2012) A small RNA that regulates motility and biofilm formation in response to changes in nutrient availability in *Escherichia coli*. *Mol. Microbiol.*, **84**, 17–35.

Tobe, T., Beatson, S. A., Taniguchi, H., Abe, H., Bailey, C. M. et al. (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.

Vanderpool, C. K. and Gottesman, S. (2004) Involvement of a novel transcriptional activator and small RNA in post-transcriptional regulation of the glucose phosphoenolpyruvate phosphotransferase system. *Mol. Microbiol.*, **54**, 1076–1089.

Vogel, J., Bartels, V., Tang, T. H., Churakov, G., SlagterJager, J. G. et al. (2003) RNomics in *Escherichia coli* detects new sRNA species and indicates parallel transcriptional output in bacteria. *Nucleic Acids Res.*, **31**, 6435–6443.

Vogel, J. and Luisi, B. F. (2011) Hfq and its constellation of RNA. *Nat. Rev. Microbiol.*, **9**, 578–589

Wassarman, K. M. (2002) Small RNAs in bacteria: Diverse regulators of gene expression in response to environmental changes. *Cell,* **109**, 141–144.

Waters, L. S. and Storz, G. (2009) Regulatory RNAs in bacteria. *Cell,* **136**, 615–628.

Zhang, A., Wassarman, K. M., Rosenow, C., Tjaden, B. C., Storz, G. et al. (2003) Global analysis of small RNA and mRNA targets of Hfq. *Mol. Microbiol.*, **50**, 1111–1124.

Zhou, X., Girón, J. A., Torres, A. G., Crawford, J. A., Negrete, E. et al. (2003) Flagellin of enteropathogenic *Escherichia coli* stimulates interleukin-8 production in T84 cells. *Infect. Immun.*, **71**, 2120–2129.

Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.,* **31**, 3406–3415.

Applied and Environmental Microbiology

Lineage-Specific Distribution of Insertion Sequence Excision Enhancer in Enterotoxigenic Escherichia coli Isolated from Swine

Masahiro Kusumoto, Dai Fukamizu, Yoshitoshi Ogura, Eiji Yoshida, Fumiko Yamamoto, Taketoshi Iwata, Tadasuke Ooka, Masato Akiba and Tetsuya Hayashi

Appl. Environ. Microbiol. **2014, 80(4):1394. DOI: 10.1128/AEM.03696-13.**
Published Ahead of Print 13 December 2013.

Updated information and services can be found at:
http://aem.asm.org/content/80/4/1394

These include:

SUPPLEMENTAL MATERIAL Supplemental material

REFERENCES This article cites 32 articles, 16 of which can be accessed free at: http://aem.asm.org/content/80/4/1394#ref-list-1

CONTENT ALERTS Receive: RSS Feeds, eTOCs, free email alerts (when new articles cite this article), more»

Information about commercial reprint orders: **http://journals.asm.org/site/misc/reprints.xhtml**
To subscribe to to another ASM Journal go to: **http://journals.asm.org/site/subscriptions/**

Journals.ASM.org

Downloaded from http://aem.asm.org/ on January 30, 2014 by NOGYO SEIBUTSU SHIGEN

AEM
Journals.ASM.org

Lineage-Specific Distribution of Insertion Sequence Excision Enhancer in Enterotoxigenic *Escherichia coli* Isolated from Swine

Masahiro Kusumoto,[a] Dai Fukamizu,[b] Yoshitoshi Ogura,[c,d] Eiji Yoshida,[e] Fumiko Yamamoto,[f] Taketoshi Iwata,[a] Tadasuke Ooka,[c] Masato Akiba,[a,g] Tetsuya Hayashi[c,d]

Bacterial and Parasitic Disease Research Division, National Institute of Animal Health, Tsukuba, Ibaraki, Japan[a]; Fukuoka Chuo Livestock Hygiene Service Center, Fukuoka, Japan[b]; Division of Microbiology, Department of Infectious Diseases, Faculty of Medicine, University of Miyazaki, Kiyotake, Miyazaki, Japan[c]; Division of Microbial Genomics, Department of Genomics and Bioenvironmental Science, Frontier Science Research Center, University of Miyazaki, Kiyotake, Miyazaki, Japan[d]; Animal Quarantine Service, Yokohama, Kanagawa, Japan[e]; Oita Livestock Hygiene Service Center, Oita, Japan[f]; Graduate School of Life and Environmental Sciences, Osaka Prefecture University, Izumisano, Osaka, Japan[g]

Downloaded from http://aem.asm.org/ on January 30, 2014 by NOGYO SEIBUTSU SHIGEN

Insertion sequences (ISs) are the simplest transposable elements and are widely distributed in bacteria; however, they also play important roles in genome evolution. We recently identified a protein called IS excision enhancer (IEE) in enterohemorrhagic *Escherichia coli* (EHEC) O157. IEE promotes the excision of IS elements belonging to the IS*3* family, such as IS*629*, as well as several other families. IEE-mediated IS excision generates various genomic deletions that lead to the diversification of the bacterial genome. IEE has been found in a broad range of bacterial species; however, among sequenced *E. coli* strains, IEE is primarily found in EHEC isolates. In this study, we investigated non-EHEC pathogenic *E. coli* strains isolated from domestic animals and found that IEE is distributed in specific lineages of enterotoxigenic *E. coli* (ETEC) strains of serotypes O139 or O149 isolated from swine. The *iee* gene is located within integrative elements that are similar to SpLE1 of EHEC O157. All *iee*-positive ETEC lineages also contained multiple copies of IS*629*, a preferred substrate of IEE, and their genomic locations varied significantly between strains, as observed in O157. These data suggest that IEE may have been transferred among EHEC and ETEC in swine via SpLE1 or SpLE1-like integrative elements. In addition, IS*629* is actively moving in the ETEC O139 and O149 genomes and, as in EHEC O157, is promoting the diversification of these genomes in combination with IEE.

Insertion sequence (IS) elements are the simplest transposable elements and are considered selfish (or parasitic) genetic elements. However, they also play important roles in genome evolution (1). The transposition and proliferation of IS elements induces not only insertional gene inactivation and the modification of gene expression (1) but also a variety of genomic rearrangements, such as deletions, inversions, and duplications (2, 3). In bacteria, several thousand types of IS elements have been identified from various species and strains (4) and classified into approximately 20 families (5).

IS-mediated bacterial genome diversification has been extensively studied in enterohemorrhagic *Escherichia coli* (EHEC) O157. EHEC O157 strains produce highly potent cytotoxins (Shiga toxins Stx1 and/or Stx2) and causes diarrhea, hemorrhagic colitis, and hemolytic-uremic syndrome; thus, it is one of the most serious food-borne infections worldwide (6). O157 strains contain many IS elements, and these elements play important roles in the diversification of the O157 genome. For example, the O157 strain RIMD0509952 (referred to as O157 Sakai) contains 25 types of IS elements (116 copies in total), and the most abundant is an IS*3* family member, IS*629* (23 copies) (7, 8). Our comparative genomic analysis of O157 clinical isolates revealed that many small structural polymorphisms associated with gene inactivation and/or deletion have been generated by IS*629* (9). More recently, we identified a novel protein called IS excision enhancer (IEE), which promotes IS*629* excision from the O157 genome in a transposase (TPase)-dependent manner. We demonstrated that various types of genomic deletions were generated upon IEE-mediated IS excision in IS-flanking regions (10). IEE also promotes the excision of other members of the IS*3*, IS*1*, and IS*30* families.

In the O157 genome, the gene encoding IEE (*iee*) is located in a large integrative element (IE) called SpLE1 (7). In non-O157 EHEC strains, *iee* is located on SpLE1-like IEs (11, 12). IEE homologs have been identified in a broad range of bacterial species and are encoded in genomic regions exhibiting low GC content and/or containing genes related to mobile genetic elements (MGEs) (10). These results suggest that IEE and its homologs have spread to a variety of bacterial strains by horizontal gene transfer. Although many *E. coli* strains have been sequenced, IEE is found primarily in EHEC isolates (10).

Pathogenic *E. coli* strains other than EHEC are also important etiological agents of zoonotic or food-borne disease in humans and of colibacillosis in domestic animals (13, 14). Enterotoxigenic *E. coli* (ETEC) is an important cause of diarrhea in children, which is associated with high morbidity and mortality in nonindustrialized countries. ETEC is the main cause of diarrhea in travelers to these countries (13). In swine, ETEC infections immediately after birth (neonatal diarrhea) and ETEC or Shiga toxin-producing *E. coli* (STEC) infections after weaning (postweaning diarrhea or edema disease) are responsible for significant economic losses due to diarrhea, growth retardation, and mortality (15, 16).

In this study, we examined the prevalence of *iee* in non-EHEC

Received 10 November 2013 Accepted 7 December 2013
Published ahead of print 13 December 2013
Address correspondence to Masahiro Kusumoto, kusu555@affrc.go.jp.
Supplemental material for this article may be found at http://dx.doi.org/10.1128/AEM.03696-13.
Copyright © 2014, American Society for Microbiology. All Rights Reserved.
doi:10.1128/AEM.03696-13

TABLE 1 Prevalence of the *iee* gene in *E. coli* isolated from domestic animals

Serotype	No. of isolates (ETEC strains)			
	Carrier animal			*iee* gene positive
	Swine	Chicken[a]	Other(s)	
O2	11 (0)	0	0	0
O7	0	1 (0)	0	0
O8	5 (4)	2 (0)	0	0
O10	0	4 (0)	0	0
O15	0	1 (0)	0	0
O16	1 (1)	0	0	0
O17	0	1 (0)	0	0
O18	0	1 (0)	0	0
O19	0	7 (0)	0	0
O25	0	1 (0)	0	0
O35	1 (1)	0	0	0
O39	0	3 (0)	0	0
O45	2 (1)	2 (0)	0	0
O56	5 (4)	0	0	0
O68	0	1 (0)	0	0
O76	0	0	3[b] (0)	0
O84	0	1 (0)	0	0
O98	5 (5)	0	0	0
O103	1 (0)	0	0	0
O115	1 (0)	0	0	0
O116	15 (15)	0	0	0
O119	0	5 (0)	0	0
O121	2 (1)	0	0	0
O123	0	1 (0)	0	0
O132	1 (0)	0	0	0
O135	0	2 (0)	0	0
O138	6 (6)	0	0	0
O139	56 (23)	0	0	23 (23)
O141	4 (2)	0	0	0
O149	17 (17)	0	0	16 (16)
O157	2 (2)	0	0	0
O161	0	1 (0)	0	0
O164	3 (3)	0	0	0
O165	0	0	1[c] (0)	0
O167	0	1 (0)	0	0
O169	1 (0)	0	0	0
Untypeable or not tested	19 (2)	59 (0)	0	0
Total	158 (87)	94 (0)	4 (0)	39 (39)

[a] These strains were all isolated from internal organs of chickens with extraintestinal diseases.
[b] These three strains were isolated from goats.
[c] This strain was isolated from cattle.

pathogenic *E. coli* strains isolated from domestic animals. Because the result indicated that the *iee* gene is distributed in specific lineages of ETEC isolated from swine, we further investigated the genomic structures of *iee*-containing IEs and the prevalence of IS elements that could be substrates for IEE in these ETEC strains.

MATERIALS AND METHODS

Bacterial strains and culture conditions. We investigated 256 *E. coli* strains, all of which were isolated from diseased domestic animals in Japan between 1991 and 2010 (Table 1). The serotypes of the strains were determined using antisera obtained from Denka Seiken Co., Ltd. (Tokyo, Japan), or Statens Serum Institut (Copenhagen, Denmark). O157 Sakai and three sequenced non-O157 EHEC strains (O26:H11 strain 11368, O111:H− strain 11128, and O103:H2 strain 12009; all three were isolated from patients in Japan [11]) were also used. All strains were grown in Luria-Bertani (LB) broth (17) at 37°C.

PCR-based genotyping of *E. coli* strains. The presence of *iee* and a panel of major IS elements (IS*629*, IS*911*, IS*3*, IS*2*, IS*1*, IS*4*, IS*5*, IS*26*, IS*30*, and IS*621*) in the *E. coli* strains was examined by PCR using the primers listed in Table S1 in the supplemental material. The template DNA for PCR was prepared by the alkaline-boiling method, as previously described (18). PCR was performed in a 50-μl reaction mixture containing template DNA, 0.2 μM concentrations of each primer, 0.2 mM concentrations of each deoxynucleoside triphosphate (dNTP), PCR buffer, and 1.25 U of ExTaq DNA polymerase (TaKaRa Bio, Inc., Shiga, Japan) using 30 amplification cycles of 30 s at 95°C, 30 s at 60°C, and 1 min at 72°C. PCR amplification of genes encoding various virulence factors (VFs; LT, STa, STb, EAST1, Stx1, Stx2, F4, F5, and F18) was performed as described by Vu-Khac et al. (19).

PFGE. Pulsed-field gel electrophoresis (PFGE) was performed by clamped homogeneous electric field electrophoresis using a CHEF-DR III apparatus (Bio-Rad Laboratories, Inc., Hercules, CA). The genomic DNA of each strain was prepared as described by Akiba et al. (20). Genomic DNA in sliced plugs was digested at 37°C with 40 U of XbaI for 6 h or 30 U of SspI for 16 h (both enzymes were obtained from TaKaRa Bio, Inc.). Electrophoresis was performed in a 1% agarose gel in 0.5× Tris-borate-EDTA (TBE) buffer at 14°C at 6 V/cm for 22 h with a pulse time of 5 to 50 s (XbaI PFGE) or for 10 h with a pulse time of 4 s (SspI PFGE). For PFGE of XbaI-digested DNA, 100 μM thiourea was added to the TBE buffer to obtain clear banding patterns (21). Southern blot hybridization analysis of SspI-digested DNA was performed using an IS*629*-specific probe, as previously described (10).

The banding patterns obtained by PFGE of XbaI-digested DNA were analyzed using BioNumerics software version 5.1 (Applied Maths, Sint-Martens-Latem, Belgium), followed by manual band assignment. Dendrograms were then generated using the unweighted pair group method with arithmetic mean (UPGMA) based on the Dice similarity index and with an optimization parameter of 1% band position tolerance.

MLST and phylogenetic tree construction. Multilocus sequence typing (MLST) was performed using the nucleotide sequences of the seven housekeeping genes *adk*, *fumC*, *gyrB*, *icd*, *mdh*, *purA*, and *recA* according to the protocols available in the *E. coli* MLST database (http://mlst.ucc.ie/mlst/mlst/dbs/Ecoli) (22). To determine the phylogenetic relationships of the *E. coli* strains, we concatenated the nucleotide sequences of the seven genes used for the MLST to generate pseudosequences and aligned them using CLUSTAL W in the software MEGA5 (23). A neighbor-joining tree was generated with a 1,000 bootstrap replicates.

Analysis of the genomic structures of *iee*-containing elements. The genomic structures of the SpLE1-like elements of the O139 and O149 strains were analyzed by PCR scanning, which is a long-range PCR-based genome comparison system (24). The primers and PCR conditions were identical to those previously described by Ohnishi et al. for whole-genome PCR scanning (WGPS) of O157 strains (24), except that the primer ieeIE-f was used instead of primer 113.3-f. To analyze the region of interest in strain E0231, we used the newly designed primer 113.9-r2 instead of 113.9-r because no amplicon was produced when the primer pair 113.8-f/113.9-r was used with this strain. In addition, to amplify both SpLE1-like IE/chromosome junctions in E0231, we designed primers IE0231-f and IE0231-r based on the results of random extension-based two-step PCR (RETS-PCR, a newly developed walking method described below). The sequences for the primers used in this analysis are shown in Table S2 in the supplemental material.

RETS-PCR. We developed a rapid walking method designated RETS-PCR to determine the sequences of the SpLE1-like IE/chromosome junctions in strain E0231. As outlined in Fig. 1, the RETS-PCR-based method comprises three steps: random extension, PCR using a single primer, and sequence determination (Fig. 1 presents the process used to analyze the left junction). In the first step, primer E0231J-R1, which comprises a

Downloaded from http://aem.asm.org/ on January 30, 2014 by NOGYO SEIBUTSU SHIGEN