



OPEN

Global distribution of epidemic-related Shiga toxin 2 encoding phages among enteroaggregative *Escherichia coli*

Keiko Kimata¹, Kenichi Lee^{2✉}, Masanori Watahiki¹, Junko Isobe¹, Makoto Ohnishi² & Sunao Lyoda²

Since the Shiga toxin-producing enteroaggregative *Escherichia coli* (Stx-EAEC) O104:H4 strain caused a massive outbreak across Europe in 2011, the importance of Stx-EAEC has attracted attention from a public health perspective. Two Stx-EAEC O86 isolates were obtained from patients with severe symptoms in Japan in 1999 and 2015. To characterize the phylogeny and pathogenic potential of these Stx-EAEC O86 isolates, whole-genome sequence analyses were performed by short- and long-read sequencing. Among genetically diverse *E. coli* O86, the Stx-EAEC O86 isolates were clustered with the EAEC O86:H27 ST3570 subgroup. Strikingly, there were only two loci with single nucleotide polymorphisms (SNPs) between the Stx2a phage of a Japanese O86:H27 isolate and that of the European epidemic-related Stx-EAEC O104:H4 isolate. These results provide evidence of global distribution of epidemic-related Stx2a phages among various lineages of *E. coli* with few mutations.

Enterotoxigenic *Escherichia coli* (EAEC) is a subgroup of diarrhoeagenic *E. coli* that shows an aggregative adherence pattern to epithelial cells¹. EAEC is a common cause of acute and persistent diarrhoea in both children and adults worldwide. EAEC infections are most frequently reported as self-limiting and result in mild symptoms²; however, Shiga toxin (Stx)-producing EAEC O104:H4 caused a massive outbreak among European countries in 2011. A total of 3,842 confirmed cases, including 845 haemolytic-uremic syndrome (HUS) cases and 54 deaths, were reported³. The epidemic showed that EAEC has the potential to become highly virulent through the acquisition of a Stx2 phage. Besides the epidemic in Europe, several cases of Stx-EAEC from HUS patients have been reported, including strains of the following serotypes: O111:H2⁴, O86:HNM⁵, O111:H21⁶, and O59:HNM⁷.

Stxs are a group of AB₅ protein toxins that inhibit protein synthesis in eukaryotic cells⁸. Shiga toxins from *E. coli* are classified into two major types: Stx1 and Stx2. These Stxs can be further subdivided into several subtypes⁹. Among the subtypes, Stx2a and Stx2d showed higher potency in cultured cells and a mouse model compared to Stx1¹⁰. Additionally, epidemiological studies have shown that *E. coli* carrying the *stx2a* operon are more likely to be involved in severe disease¹¹. The *stx* genes are encoded in Lambda-like lysogenic phages. Therefore, Stx-encoding phages can be transferred horizontally, and phages have been found in various lineages of *E. coli* though rarely in other Enterobacteriaceae^{12,13}.

A subset of *E. coli* producing Stx (or carrying the *stx* gene) is called Shiga toxin-producing *E. coli* (STEC). STEC are important foodborne pathogens that cause enteritis, bloody diarrhoea, and often fatal HUS worldwide¹⁴. As more than 3,000 cases of infection have been reported annually in Japan¹⁵, national surveillance has been performed by molecular typing methods, including multilocus variable-number tandem repeat analysis, pulsed field gel electrophoresis, and whole-genome sequencing (WGS), since 1996^{16–19}. From this surveillance, two isolates of Stx-EAEC serogroup O86 from bloody diarrhoea or HUS, and two STEC O86 isolates from asymptomatic carriers were reported by 2017; however, there is no known epidemiological link or genetic information shared between the isolates. In this study, to elucidate the genomic characteristics and phylogenetic lineage of

¹Department of Bacteriology, Toyama Institute of Health, Toyama 9390363, Japan. ²Department of Bacteriology I, National Institute of Infectious Diseases, Tokyo 1628640, Japan. ✉email: leek@niid.go.jp

ID	BioSample	Serotype ^a	Year	Symptom	Pathotype	Phylogenetic group	<i>stx</i>
JE86-ST02	SAMD00137267	O86:H27	1999	HUS	Stx-EAEC	B1	2a
JE86-ST03	SAMD00137268	O86:H32	2007	Asymptomatic	STEC	A	2e
JE86-ST04	SAMD00137269	O86:H51	2011	Asymptomatic	STEC	B1	1a
JE86-ST05	SAMD00137270	O86:H27	2014	Bloody diarrhea	Stx-EAEC	B1	2a

Table 1. Strains used in this study. ^aSerotype predicted by the draft genomes. All the isolate showed non-motile phenotype.

STEC O86 isolated in Japan in the last 20 years, including two Stx-EAEC, the draft genomes were analysed and compared with published WGS data of various *E. coli* strains.

To obtain the information of the most important virulence determinants, Stx2a phages and virulence plasmids, the complete genomes of Stx-EAEC O86 isolates were determined by long read sequencing. Comparative genome analyses revealed that the Stx2a phage of one Stx-EAEC O86 isolate was nearly identical to that of the European Stx-EAEC O104:H4 strain, suggesting the global distribution of the epidemic-related Stx2a phage.

Results

Characterization of phylogeny and pathogenic potential of Japanese STEC O86 isolates by WGS analyses using draft genome sequences. From 1999 to 2017, four STEC O86 isolates were detected by national surveillance (Table 1). To infer the phylogenetic relationships and characterize the virulence profiles of the isolates, WGS analyses of these isolates was performed and data from public databases were analysed. WGS data from a public database (Enterobase, <https://enterobase.warwick.ac.uk/species/index/ecoli>) were collected from international isolates of *E. coli* O86, Stx-EAEC, and other pathogenic or commensal *E. coli* strains derived from humans and animals, as shown in Supplementary Table S1. At the time of writing, three Stx-EAEC isolates, other than O86 and O104:H4, had been reported to cause HUS as follows: O111:H2⁴, O111:H21⁶, and O59:HNM⁷. Among them, a draft genome of an O111:H21 isolate from the UK was available. This isolate was used for phylogenetic analysis and virulence factor comparison. For O111:H2 isolates from France, only the Stx2 phage sequence was available, and this was used for phage comparison as described below.

Although all the STEC O86 in Japan was non-motile, sequencing of the flagellar gene, *fliC*, identified three H-types (H27, H32, and H51). Notably, these serotypes can be further subdivided by the sequence type (ST) by multilocus sequence typing (MLST) (Fig. 1, complete information is shown in Supplementary Table S2). Core genome single nucleotide polymorphism (cgSNP)-based phylogenetic analyses reinforced the result that each serotype includes different lineages. The Japanese O86 isolates could be divided into the following three distinct subgroups: Stx-EAEC O86:H27 ST3570 (JE86-ST02 and 05), STEC O86:H51 ST155 (JE86-ST04), and STEC O86:H32 ST5133 (JE86-ST03). In the first subgroup, two isolates of Stx-EAEC O86:H27 were clustered together with other EAEC O86:H27; however, the other EAEC O86:H27 isolates did not carry *stx2*, except EH3148 (from Belgium in 2018). In the second subgroup (O86:H51 ST155), one Japanese isolate (JE86-ST04) was clustered with three isolates from Japan and the USA. JE86-ST04 possessed the *stx1a* and STEC autoagglutinating adhesin (Saa) gene, which is one of adhesins of locus of enterocyte effacement (LEE)-negative STEC²⁰. The other isolates in this cluster did not carry these genes, suggesting that JE86-ST04 could have emerged following their acquisition of the gene in an *stx*-negative O86:H51 strain. The third subgroup was *stx2e*-positive O86:H32 ST5133. The isolates in this subgroup, except FSIS11815123, possessed *stx2e*. *Stx2e* is a major virulence factor for porcine edema disease²¹ and is rarely associated with human HUS²². Although the principal adhesin of edema disease-causing *E. coli*, F18 fimbriae²³; was not detected, another adhesin for porcine *E. coli*, AIDA (adhesin involved in diffuse adherence)²³ was detected.

A number of key components involved in the pathogenesis of EAEC have been reported, although not fully understood. The presence and similarity of these virulence factors were compared using draft genomes of EAEC O86:H27 isolates and previously sequenced Stx-EAEC isolates (Supplementary Table S2). The sequences of virulence regulator gene of EAEC (*aggR*), *aggR*-regulated dispersin gene (*aap*), dispersin transporter operon (*aat*), and type 6 secretion system operon (*aai*), were highly similar (>95%) among the isolates. In contrast, although all the isolates possessed aggregative adherence fimbriae (AAF), they showed high diversity among the serotypes: O86:H27, AAF/III; O104:H4, AAF/I; O111:H21, AAF/V (Supplementary Table S2).

Complete genome sequences of Stx-EAEC O86:H27. By using PacBio and Oxford Nanopore sequencing along with Illumina short read sequencing, complete genome sequences of the Stx-EAEC O86:H27 isolates were determined. Genome statistics are shown in Table 2. Both isolates had a 5.3 Mb chromosome, which is much larger than *E. coli* K12 (ca. 4.7 Mb). PHASTER analysis showed that JE86-ST02 and JE86-ST05 harboured eight and ten putative prophage sequences, respectively. They both had a large plasmid (115 kb) that harboured pAA-borne virulence genes, including *aggR*, an *aaf* gene cluster, *aat*, and an *aap* operon (Fig. 2). Both isolates carried the β -lactamase gene, *bla*_{TEM-1B}, in the chromosome, while the epidemic Stx-EAEC O104:H4 strain harbors extended-spectrum β -lactamase genes, including *bla*_{CTX-M-15} and *bla*_{TEM-1B} in an 89 kb- plasmid.

Comparative genomics of Stx2a phages and plasmids of Stx-EAEC O86:H27. When the complete sequences of the Stx2a phages of JE86-ST02 and JE86-ST05 were compared, only some CDSs in the late

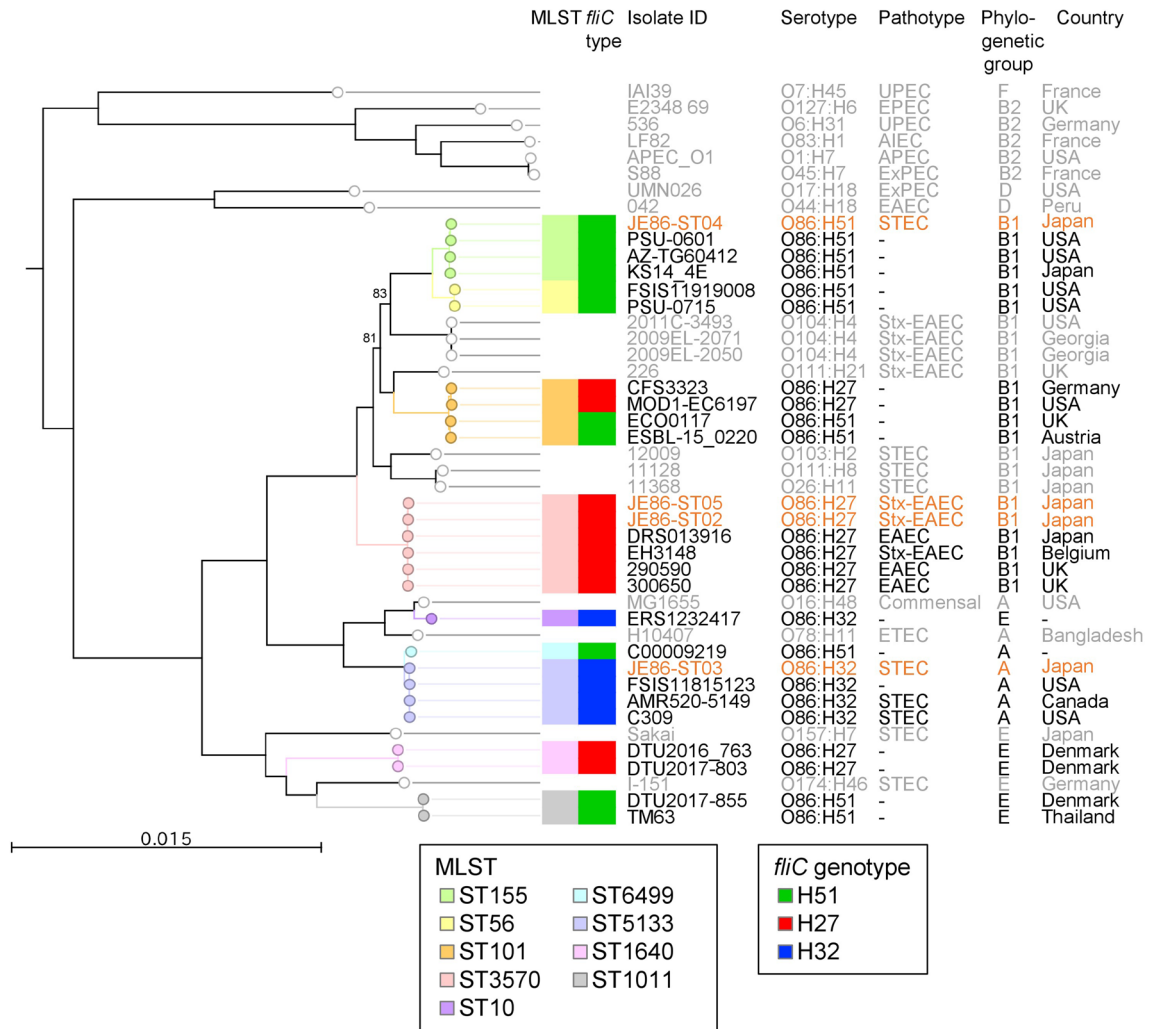


Figure 1. Core genome SNP-based maximum likelihood tree of *Escherichia coli* O86 and other *E. coli* serogroups. The colour of branches and nodes (open circles) represents the sequence type (ST) by multilocus sequence typing (MLST). The colour of boxes on the right represents the ST and the *fliC* genotype (in silico H type). Isolate names sequenced in this study are shown in red. Information of the non-O86 isolates was shaded. The tree was rooted by *E. albertii* 2012EL-1823B. Bootstrap values below 95 are shown at the branch. Abbreviations of the pathotypes were as follows: UPEC, uropathogenic *E. coli*; EPEC, enteropathogenic *E. coli*; AIEC, adherent invasive *E. coli*; APEC, avian pathogenic *E. coli*; ExPEC, extraintestinal pathogenic *E. coli*; EAEC, enteroaggregative *E. coli*; STEC, Shiga toxin-producing *E. coli*; Stx-EAEC, Shiga toxin-producing enteroaggregative *E. coli*; ETEC, enterotoxigenic *E. coli*; -, not available. The phylogenetic group was determined by in silico PCR.

region showed similarity (80–99%) (Fig. 3). The integrases of the phages were different and the insertion site of the phages was different (JE86-ST02, *argW*; JE86-ST05, *wrbA*). Interestingly, the results of NCBI-BLAST (<https://blast.ncbi.nlm.nih.gov/>) showed that the Stx2a phage of JE86-ST05 was highly similar to that of the European epidemic O104:H4 strains (Supplementary Table S4). There were only two loci of SNPs between the phage of JE86-ST05 and that of the European epidemic strain, 2011C-3493 (Fig. 3, Supplementary Table S4). A similar phage from Stx-EAEC O111:H2 (strain ED191, accession no. KF971864) has been reported previously²⁴. This strain was the causative agent of the French outbreak in 1992^{24,25}. There were 56 SNPs and 378 indels between the O104:H4 2011C-3493 and O111:H2 ED191 phages (Fig. 3). It is of note that the similar phages have been found from other *E. coli* pathotypes. The Stx2a phage from hybrid Shigatoxigenic and enterotoxigenic *E. coli* (STEC/ETEC) isolated from Finland in 2001²⁶ had only 45 bp-deletion compared to that of O104:H4 2011C-3493 (Fig. 3). Additionally, the phage of STEC O26 isolates from patients in Europe showed the same phage structure to that of O104:H4 2011C-3493²⁷. Meanwhile, similar phages to the Stx2a phage of JE86-ST02 were not found in public databases. Stx2a phages were also found in Stx-EAEC O111:H21 226 from Northern Ireland and Stx-EAEC O86:H27 EH3148 from Belgium. However, their Stx2a phage sequences were quite distinct from both JE86-ST02 and JE86-ST05, according to the draft genome analysis (data not shown).

Statistics	JE86-ST02				JE86-ST05					
	Chromosome	p1	p2	p3	Chromosome	p1	p2	p3	p4	p5
Size (kb)	5,283,470	114,901	83,276	48,397	5,327,513	114,953	85,841	48,666	7,939	5,716
No. CDS	4,996	132	93	71	5,103	134	99	69	9	6
No. rRNA	22	0	0	0	22	0	0	0	0	0
No. tRNA	105	0	0	1	98	0	0	1	0	0
No. prophage-like elements	8	0	0	0	10	0	0	1	0	0
GC%	50.7	48.7	53.1	44.6	50.8	48.7	53.1	44.5	41.6	42.4
Chromosomal sequence type or plasmid Inc type	ST3570	IncFIB, IncFII	IncB/O/K/Z	ND	ST3570	IncFIB, IncFII	IncB/O/K/Z	ND	ND	ND
Antimicrobial resistance gene(s)	<i>aadA1</i> , <i>ant(3'')-Ia</i> , <i>blaTEM-1B</i> , <i>mdf(A)</i> , <i>sul1</i>	ND	ND	ND	<i>blaTEM-1B</i> , <i>mdf(A)</i>	ND	ND	ND	ND	ND
Accession no	AP022811	AP022812	AP022813	AP022814	AP022815	AP022816	AP022817	AP022818	AP022819	AP022820

Table 2. Statistics of the complete genome sequence. *ND* not detected.

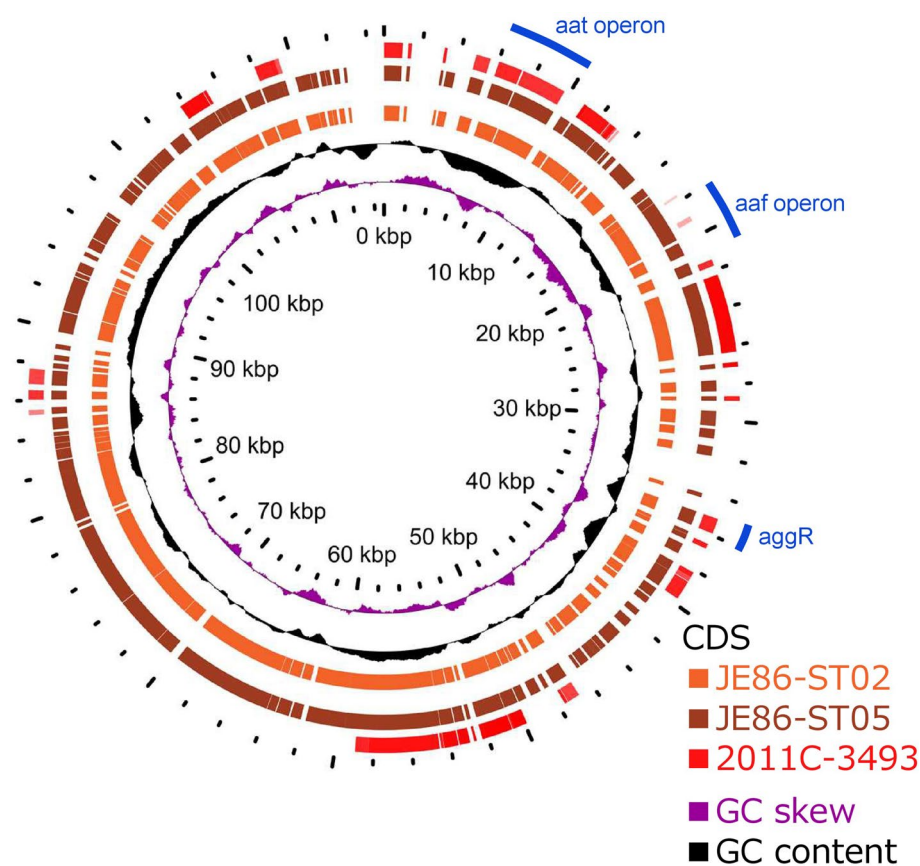


Figure 2. Comparison of the virulence plasmid (pAA) of three isolates of Shiga toxin-producing enteroaggregative *E. coli*. BLAST Atlas analysis was performed by GView Server. A lane for JE86-ST02 represents coding sequences (CDSs) of the pAA. Lanes for JE86-ST05 and 2011C-3493 represent CDSs of pAA that showed more than 80% similarity to the plasmid of JE86-ST02.

In contrast, the pAA plasmids of Stx-EAEC O86:H27 are highly similar to each other, but they had low similarity to that of O104:H4 2011C-3493 (Fig. 2). All the coding sequences (CDSs) of JE86-ST02 and JE86-ST05 were shared. The mean \pm SD similarity of all CDSs between the two plasmids was $99.9 \pm 1.3\%$.

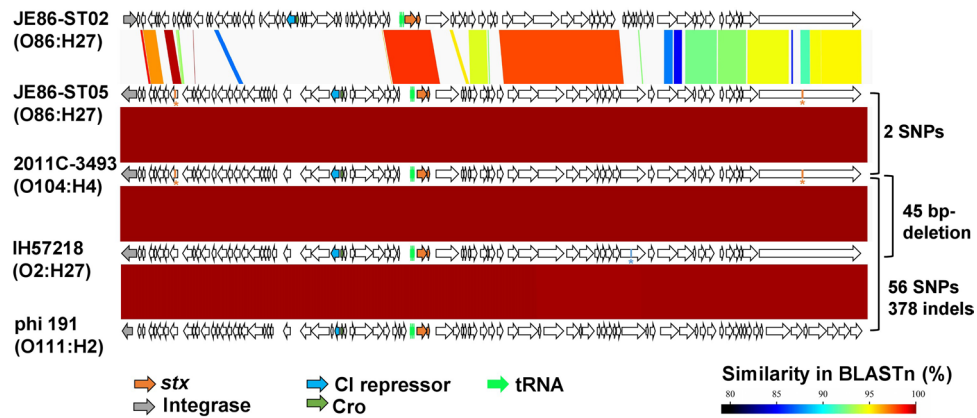


Figure 3. Comparison of the complete sequence of the Shiga toxin (Stx) 2a phage of Stx-enteroaggregative *E. coli* O86:H27, O104:H4, hybrid Shigatoxigenic and enterotoxigenic *E. coli* O2:H27, and O111:H2. Arrows represent the CDSs, and major genes are coloured as shown in the legend. BLASTn and visualization were performed with GenomeMatcher software. The number of single nucleotide polymorphisms (SNPs) and insertions and deletions (indels) to the phage of O104:H4 are shown on the right. Red and blue asterisks show the position of SNPs and deletion, respectively, compared to the phage of O104:H4.

Discussion

In this study, we revealed the genomic diversity of domestic and international *E. coli* O86 and found evidence of global distribution of the epidemic-related Stx2a phages.

According to WGS-based phylogeny, the polyphyletic nature of *E. coli* O86 was shown. Serogroup O86 consists of three serotypes (O86:H27, H32, and H51), and each serotype was found in multiple lineages. Even isolates showing different H type clustered together (i.e. in ST101 includes H27 and H51 isolates). This is not surprising because the O- and H-antigen gene clusters can be mobile via recombination^{28–30}. Domestic O86 isolates could be divided into three distinct subgroups. Stx-EAEC O86 isolates were clustered with EAEC O86:H27 from Belgium, the UK, and Japan. These isolates shared almost identical virulence components of EAEC, including *aggR*, *aap*, the *aat* operon, T6SS, and SPATEs (Supplementary Table S3), suggesting closely-related EAEC is distributed in Europe and Japan. Although the symptoms caused by the other EAEC O86:H27 isolates were not available, Stx2a might increase the virulence of the bacterium. The other two isolates, JE86-ST03 and 04, were obtained from asymptomatic carriers and thus, their pathogenicity was unclear due to low sample size. JE86-ST03 and JE86-ST04 carried several putative virulence factors of diarrheagenic *E. coli*, including AIDA, *saa*, *cdt*, and *espP*. Further studies are required to elucidate whether these factors may have virulence potential to humans.

Notably, one of the Stx-EAEC O86:H27 isolates harboured an Stx2a phage identical to that of European epidemic Stx-EAEC O104:H4 isolates. A highly similar phage was found in Stx-EAEC O111:H2 19 years before the European epidemic caused by O104:H4^{24,25}. The related phages have been found in STEC/ETEC O2:H27 (Finland) and STECs O26 (Europe), providing the evidence of global-level distribution of the Stx2a phage. On the other hand, the two genetically close isolates of Stx-EAEC O86:H27 harboured two distinct Stx2a phages. The two Stx-EAEC isolates, O111:H21 226 from Northern Ireland⁶ and O86:H27 EH3148 from Belgium, also harboured a different Stx2a phage from the other Stx-EAEC. These results provide two implications. First, the epidemic related Stx2a phage is globally distributed among various *E. coli* pathotypes. Second, EAEC has the potential to acquire different Stx2 phages. It is likely that global transfer of the host bacterium facilitates the phage distribution. The principal infection route of EAEC is believed to be human-to-human transmission. Meanwhile, food and animals are indicated to be involved in the transmission of Stx-EAEC O104:H4 and STEC O26³¹, and the related phages were found from cattle³². Therefore, various source and transmission routes should be considered to understand the epidemiology of the Stx phage and emergence of highly virulent Stx-EAEC.

Regarding the pathogenicity of EAEC, the pAA and *aggR*-regulated genes have been regarded as key virulence factors in EAEC in vivo and in vitro settings^{33,34}, while other report emphasizes the importance of certain SPATEs encoded in the chromosome or plasmids³⁵. Both Stx-EAEC O86:H27 isolates possessed a set of virulence factors, including, *aggR*, *aggR* regulons, some SPATEs, and other toxins, as other EAEC strains do. Therefore, severe symptoms (e.g. bloody diarrhoea or HUS) by Stx-EAEC O86:H27 may be explained by Stx2a production.

In conclusion, our results indicate the potential of the Stx2a phage of Stx-EAEC O104:H4 to transfer horizontally into phylogenetically distinct strains with few mutations. It is plausible that the Stx2a phage is circulating among various pathotypes of *E. coli*. To understand the epidemiology of the phage in detail, STEC from various locations, sources, pathotypes, and lineages should be considered.

Methods

Isolates used in this study. From 1999 to 2017, four STEC O86 isolates were reported through national surveillance (Table 1). Two isolates were from patients, while two isolates were obtained from asymptomatic carriers. These isolates were originally isolated by local health institutes and subsequently sent to our laboratory for further analysis. JE86-ST02 corresponds to 990599 in a previous report by Iyoda et al.⁵

WGS phylogeny and in silico typing of draft genomes. Genomic DNA was extracted with the DNeasy Blood & Tissue Kit (QIAGEN, Venlo, Netherlands) and Genomic-tip 100/G (QIAGEN) for short-read and long-read sequencing, respectively. For short-read sequencing, genomic DNA libraries were prepared using a Nextera XT DNA Sample Prep Kit (Illumina, San Diego, CA, USA). The pooled libraries were subjected to multiplexed paired-end sequencing (300-mer \times 2) using MiSeq (Illumina). The short reads were assembled using SPAdes v.3.11.1 with the “-careful” option³⁶. Contigs of each isolate were comprehensively characterized using an in-house BLAST-based pipeline as described previously¹⁸. Phylogenetic relationships of STEC O86 isolates and non-O86 *E. coli* belonging to various pathotypes and phylogenetic groups (Supplementary Table S1) were inferred by mapping-based analysis. Core genome SNPs were extracted by using BactSNP v.1.1.0³⁷ with the genome of STEC O157 strain Sakai (GenBank accession No.: BA000007) as a reference. Repetitive regions longer than 50 bp were detected by MUMmer v.3.2259 (nucmer, repeat-match, and exact-tandems functions)³⁸ and removed for further analyses, as were prophage regions. The recombinogenic regions were detected by gubbins³⁹ and removed. Finally, 244,878 SNP sites from 2,074,586 bp of conserved backbone were used for further analyses. Phylogenetic relationships were determined by reconstructing a phylogenetic tree using the maximum likelihood method using IQ-TREE with 1,000 ultrafast bootstrap replicates⁴⁰.

Determination of the complete genome sequence of Stx-EAEC O86:H27

For the O86:H27 isolates, complete genome sequences were determined by using long read sequencing. For JE86-ST02, PacBio RS II (Pacific Biosciences, Menlo Park, CA) sequencing was performed using the PacBio SMRTbell Template Prep Kit 1.0 and Polymerase Binding Kit P6 after size selection using BluePippin (Sage Science, Beverly, MA) with a cutoff value of 20 kb. The long-read sequences were assembled using the Hierarchical Genome Assembly Process (HGAP) version 3 with SMRT Analysis software⁴¹ and polished by Pilon⁴² with short-read sequences. For JE86-ST05, a sequencing library was prepared by using a Rapid Barcoding Sequencing Kit (SQK-RBK004, Oxford Nanopore Technologies, Oxford, UK). A MinION R9.4 flow cell (Oxford Nanopore) was used for 48 h sequencing. The long-read sequences and the short-read sequences were subjected to hybrid assembly by using Unicycler v.0.4.4⁴³. Annotation for the complete genomes was performed by DFAST⁴⁴ with manual curation. Prophage sequences in the genome were detected by PHASTER⁴⁵. The complete sequences of the pAA plasmids of the O86:H27 isolates were compared to Stx-EAEC O104:H4 2011C-3493 by using GView Server (<https://server.gview.ca/>). The Stx2a phage sequences of the three isolates used in the plasmid comparison and strain O111:H2 ED191 and O2:H27 IH57218 were compared and visualized by using GenomeMatcher⁴⁶.

Ethics statement. Patient information of the isolates was completely anonymized.

Data availability

The complete-genome sequences of JE86-ST02 and JE86-ST05 were submitted to DDBJ/ENA/GenBank under the accession number AP022811-AP022820 as shown in Table 2. Short reads and draft genomes of Japanese *E. coli* O86 isolates were deposited under the accession number DRA007294.

Received: 26 February 2020; Accepted: 8 June 2020

Published online: 16 July 2020

References

- Huang, D. B., Mohanty, A., DuPont, H. L., Okhuysen, P. C. & Chiang, T. A review of an emerging enteric pathogen: enteroaggregative *Escherichia coli*. *J. Med. Microbiol.* **55**, 1303–1311. <https://doi.org/10.1099/jmm.0.46674-0> (2006).
- Hebbelstrup Jensen, B., Olsen, K. E., Struve, C., Krogfelt, K. A. & Petersen, A. M. Epidemiology and clinical manifestations of enteroaggregative *Escherichia coli*. *Clin. Microbiol. Rev.* **27**, 614–630. <https://doi.org/10.1128/CMR.00112-13> (2014).
- Frank, C. *et al.* Epidemic profile of Shiga-toxin-producing *Escherichia coli* O104:H4 outbreak in Germany. *N. Engl. J. Med.* **365**, 1771–1780. <https://doi.org/10.1056/NEJMoa1106483> (2011).
- Morabito, S. *et al.* Enteroaggregative, Shiga toxin-producing *Escherichia coli* O111:H2 associated with an outbreak of hemolytic-uremic syndrome. *J. Clin. Microbiol.* **36**, 840–842 (1998).
- Iyoda, S. *et al.* Inducible *stx2* phages are lysogenized in the enteroaggregative and other phenotypic *Escherichia coli* O86:HNM isolated from patients. *FEMS Microbiol. Lett.* **191**, 7–10. <https://doi.org/10.1111/j.1574-6968.2000.tb09311.x> (2000).
- Dallman, T. *et al.* Characterization of a verocytotoxin-producing enteroaggregative *Escherichia coli* serogroup O111:H21 strain associated with a household outbreak in Northern Ireland. *J. Clin. Microbiol.* **50**, 4116–4119. <https://doi.org/10.1128/JCM.02047-12> (2012).
- Carbonari, C. C. *et al.* An Stx-EAEC O59:NM[H19] strain isolated from a hemolytic uremic syndrome case in Argentina. *Rev. Argent. Microbiol.* <https://doi.org/10.1016/j.ram.2019.03.006> (2019).
- Lee, M. S., Koo, S., Jeong, D. G. & Tesh, V. L. Shiga toxins as multi-functional proteins: induction of host cellular stress responses, role in pathogenesis and therapeutic applications. *Toxins*. <https://doi.org/10.3390/toxins8030077> (2016).
- Scheutz, F. *et al.* Multicenter evaluation of a sequence-based protocol for subtyping Shiga toxins and standardizing Stx nomenclature. *J. Clin. Microbiol.* **50**, 2951–2963. <https://doi.org/10.1128/JCM.00860-12> (2012).
- Fuller, C. A., Pellino, C. A., Flagler, M. J., Strasser, J. E. & Weiss, A. A. Shiga toxin subtypes display dramatic differences in potency. *Infect. Immun.* **79**, 1329–1337. <https://doi.org/10.1128/IAI.01182-10> (2011).
- Melton-Celsa, A. R. Shiga toxin (Stx) classification, structure, and function. *Microbiol. Spectr.* <https://doi.org/10.1128/microbiolspec.EHEC-0024-2013> (2014).
- Khalil, R. K., Skinner, C., Patfield, S. & He, X. Phage-mediated Shiga toxin (Stx) horizontal gene transfer and expression in non-Shiga toxigenic *Enterobacter* and *Escherichia coli* strains. *Pathog. Dis.* <https://doi.org/10.1093/femspd/ftw037> (2016).
- Muniesa, M. *et al.* Shiga toxin 2-converting bacteriophages associated with clonal variability in *Escherichia coli* O157:H7 strains of human origin isolated from a single outbreak. *Infect. Immun.* **71**, 4554–4562. <https://doi.org/10.1128/iai.71.8.4554-4562.2003> (2003).
- Pennington, H. *Escherichia coli* O157. *Lancet* **376**, 1428–1435. [https://doi.org/10.1016/S0140-6736\(10\)60963-4](https://doi.org/10.1016/S0140-6736(10)60963-4) (2010).

15. Infectious Agents Surveillance Report. *Enterohemorrhagic Escherichia coli (EHEC) infection, as of March 2018, Japan*. <https://www.niid.go.jp/niid/en/iasr-vol39-e/865-iasr/8042-459te.html> (2018).
16. Izumiya, H. *et al.* New system for multilocus variable-number tandem-repeat analysis of the enterohemorrhagic *Escherichia coli* strains belonging to three major serogroups: O157, O26, and O111. *Microbiol. Immunol.* **54**, 569–577. <https://doi.org/10.1111/j.1348-0421.2010.00252.x> (2010).
17. Swaminathan, B. *et al.* Building PulseNet International: an interconnected system of laboratory networks to facilitate timely public health recognition and response to foodborne disease outbreaks and emerging foodborne diseases. *Foodborne Pathog. Dis.* **3**, 36–50. <https://doi.org/10.1089/fpd.2006.3.36> (2006).
18. Lee, K., Izumiya, H., Iyoda, S. & Ohnishi, M. Effective surveillance using multilocus variable-number tandem-repeat analysis and whole-genome sequencing for enterohemorrhagic *Escherichia coli* O157. *Appl. Environ. Microbiol.* <https://doi.org/10.1128/AEM.00728-19> (2019).
19. Lee, K. *et al.* A geographically widespread outbreak investigation and development of a rapid screening method using whole genome sequences of enterohemorrhagic *Escherichia coli* O121. *Front. Microbiol.* **8**, 701. <https://doi.org/10.3389/fmicb.2017.00701> (2017).
20. Paton, A. W., Srimanote, P., Woodrow, M. C. & Paton, J. C. Characterization of Saa, a novel autoagglutinating adhesin produced by locus of enterocyte effacement-negative Shiga-toxicogenic *Escherichia coli* strains that are virulent for humans. *Infect. Immun.* **69**, 6999–7009. <https://doi.org/10.1128/IAI.69.11.6999-7009.2001> (2001).
21. Casanova, N. A., Redondo, L. M., Dailoff, G. C., Arenas, D. & Fernandez Miyakawa, M. E. Overview of the role of Shiga toxins in porcine edema disease pathogenesis. *Toxicon* **148**, 149–154. <https://doi.org/10.1016/j.toxicon.2018.04.019> (2018).
22. Fasel, D. *et al.* Hemolytic uremic syndrome in a 65-Year-old male linked to a very unusual type of *stx2e*- and *eae*-harboring O51:H49 Shiga toxin-producing *Escherichia coli*. *J. Clin. Microbiol.* **52**, 1301–1303. <https://doi.org/10.1128/JCM.03459-13> (2014).
23. Niewerth, U. *et al.* The AIDA autotransporter system is associated with F18 and *stx2e* in *Escherichia coli* isolates from pigs diagnosed with edema disease and postweaning diarrhea. *Clin. Diagn. Lab. Immunol.* **8**, 143–149. <https://doi.org/10.1128/cdli.8.1.143-149.2001> (2001).
24. Grande, L. *et al.* Whole genome sequence comparison of *vtx2*-converting phages from enteroaggregative haemorrhagic *Escherichia coli* strains. *BMC Genom.* **15**, 574. <https://doi.org/10.1186/1471-2164-15-574> (2014).
25. Boudailliez, B. *et al.* Possible person-to-person transmission of *Escherichia coli* O111-associated hemolytic uremic syndrome. *Pediatr. Nephrol.* **11**, 36–39. <https://doi.org/10.1007/s004670050229> (1997).
26. Nyholm, O. *et al.* Comparative genomics and characterization of hybrid Shiga-toxicogenic and enterotoxigenic *Escherichia coli* (STEC/ETEC) strains. *PLoS ONE* **10**, e0135936. <https://doi.org/10.1371/journal.pone.0135936> (2015).
27. Karnisova, L. *et al.* Attack of the clones: whole genome-based characterization of two closely related enterohemorrhagic *Escherichia coli* O26 epidemic lineages. *BMC Genom.* **19**, 647. <https://doi.org/10.1186/s12864-018-5045-7> (2018).
28. Geue, L. *et al.* Evidence for contemporary switching of the O-antigen gene cluster between Shiga toxin-producing *Escherichia coli* strains colonizing cattle. *Front. Microbiol.* **8**, 424. <https://doi.org/10.3389/fmicb.2017.00424> (2017).
29. Sugiyama, T. *et al.* Generation of *Escherichia coli* O9a serotype, a subtype of *E. coli* O9, by transfer of the *wb** gene cluster of *Klebsiella* O3 into *E. coli* via recombination. *J. Bacteriol.* **180**, 2775–2778 (1998).
30. Wang, L., Rothmund, D., Curd, H. & Reeves, P. R. Species-wide variation in the *Escherichia coli* flagellin (H-antigen) gene. *J. Bacteriol.* **185**, 2936–2943. <https://doi.org/10.1128/jb.185.9.2936-2943.2003> (2003).
31. Navarro-Garcia, F. *Escherichia coli* O104:H4 pathogenesis: an enteroaggregative *E. coli*/Shiga toxin-producing *E. coli* explosive cocktail of high virulence. *Microbiol. Spectr.* <https://doi.org/10.1128/microbiolspec.EHEC-0008-2013> (2014).
32. Beutin, L., Hammer, J. A., Reetz, J. & Strauch, E. Shiga toxin-producing *Escherichia coli* strains from cattle as a source of the Stx2a bacteriophages present in enteroaggregative *Escherichia coli* O104:H4 strains. *Int. J. Med. Microbiol.* **303**, 595–602. <https://doi.org/10.1016/j.ijmm.2013.08.001> (2013).
33. Boisen, N. *et al.* The presence of the pAA plasmid in the German O104:H4 Shiga toxin type 2a (Stx2a)-producing enteroaggregative *Escherichia coli* strain promotes the translocation of Stx2a across an epithelial cell monolayer. *J. Infect. Dis.* **210**, 1909–1919. <https://doi.org/10.1093/infdis/jiu399> (2014).
34. Boisen, N. *et al.* The role of the AggR regulon in the virulence of the Shiga toxin-producing enteroaggregative *Escherichia coli* epidemic O104:H4 strain in mice. *Front. Microbiol.* **10**, 1824. <https://doi.org/10.3389/fmicb.2019.01824> (2019).
35. Munera, D. *et al.* Autotransporters but not pAA are critical for rabbit colonization by Shiga toxin-producing *Escherichia coli* O104:H4. *Nat. Commun.* **5**, 3080. <https://doi.org/10.1038/ncomms4080> (2014).
36. Bankevich, A. *et al.* SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **19**, 455–477. <https://doi.org/10.1089/cmb.2012.0021> (2012).
37. Yoshimura, D. *et al.* Evaluation of SNP calling methods for closely related bacterial isolates and a novel high-accuracy pipeline: BactSNP. *Microb. Genom.* <https://doi.org/10.1099/mgen.0.000261> (2019).
38. Kurtz, S. *et al.* Versatile and open software for comparing large genomes. *Genome Biol.* **5**, R12. <https://doi.org/10.1186/gb-2004-5-2-r12> (2004).
39. Croucher, N. J. *et al.* Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. *Nucleic Acids Res.* **43**, e15. <https://doi.org/10.1093/nar/gku1196> (2015).
40. Nguyen, L.-T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274. <https://doi.org/10.1093/molbev/msu300> (2015).
41. Chin, C. S. *et al.* Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat. Methods* **10**, 563–569. <https://doi.org/10.1038/nmeth.2474> (2013).
42. Walker, B. J. *et al.* Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS ONE* **9**, e112963. <https://doi.org/10.1371/journal.pone.0112963> (2014).
43. Wick, R. R., Judd, L. M., Gorrie, C. L. & Holt, K. E. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput. Biol.* **13**, e1005595. <https://doi.org/10.1371/journal.pcbi.1005595> (2017).
44. Tanizawa, Y., Fujisawa, T. & Nakamura, Y. DFAST: a flexible prokaryotic genome annotation pipeline for faster genome publication. *Bioinformatics* **34**, 1037–1039. <https://doi.org/10.1093/bioinformatics/btx713> (2018).
45. Arndt, D. *et al.* PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res.* **44**, W16–W21. <https://doi.org/10.1093/nar/gkw387> (2016).
46. Ohtsubo, Y., Ikeda-Ohtsubo, W., Nagata, Y. & Tsuda, M. GenomeMatcher: a graphical user interface for DNA sequence comparison. *BMC Bioinform.* **9**, 376 (2008).

Acknowledgements

JE-ST86-05 was provided by the Department of Microbiology, Miyagi Prefectural Institute of Public Health and Environment. We are grateful to Saomi Ozawa, Yukie Nakajima, Ayumi Takemoto, and Yu Takizawa for technical assistance and WGS analyses. This work was partially supported by the Research Program on Emerging and Re-emerging Infectious Diseases from the Japan Agency for Medical Research and Development (AMED) (Grant Number JP19fk0108065) and JSPS KAKENHI (Grant Number 18K15156).

Author contributions

K.K., K.L., M.O., and S.I. conceived and designed the experiments. K.K., K.L., M.W., and J.I. performed the experiments and analysed the data. K.K., K.L., M.W., J.I., M.O., and S.I. wrote the paper.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41598-020-68462-9>.

Correspondence and requests for materials should be addressed to K.L.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020