AEM
Journals.ASM.org

Sequence Variability of P2-Like Prophage Genomes Carrying the Cytolethal Distending Toxin V Operon in *Escherichia coli* O157

Domonkos Sváb,[a] Balázs Horváth,[a] Gergely Maróti,[b] Ulrich Dobrindt,[c] István Tóth[a]

Institute for Veterinary Medical Research, Centre for Agricultural Research, Hungarian Academy of Sciences, Budapest, Hungary[a]; Institute of Biochemistry, Biological Research Centre, Hungarian Academy of Sciences, Szeged, Hungary[b]; Institute of Hygiene, Westfalian Wilhelms University Münster, Münster, Germany[c]

Cytolethal distending toxins (CDT) are potent cytotoxins of several Gram-negative pathogenic bacteria, including *Escherichia coli*, in which five types (CDT-I to CDT-V) have been identified so far. CDT-V is frequently associated with Shiga-toxigenic *E. coli* (STEC), enterohemorrhagic *E. coli* (EHEC) O157 strains, and strains not fitting any established pathotypes. In this study, we were the first to sequence and annotate a 31.2-kb-long, noninducible P2-like prophage carrying the *cdt-V* operon from an *stx*- and *eae*-negative *E. coli* O157:H43 strain of bovine origin. The *cdt-V* operon is integrated in the place of the *tin* and *old* phage immunity genes (termed the TO region) of the prophage, and the prophage itself is integrated into the bacterial chromosome between the housekeeping genes *cpxP* and *fieF*. The presence of P2-like genes ($n = 20$) was investigated in a further five CDT-V-positive bovine *E. coli* O157 strains of various serotypes, three EHEC O157:NM strains, four strains expressing other variants of CDT, and eight CDT-negative strains. All but one CDT-V-positive atypical O157 strain uniformly carried all the investigated genomic regions of P2-like phages, while the EHEC O157 strains missed three regions and the CDT-V-negative strains carried only a few P2-like sequences. Our results suggest that P2-like phages play a role in the dissemination of *cdt-V* between *E. coli* O157 strains and that after integration into the bacterial chromosome, they adapted to the respective hosts and became temperate.

Cytolethal distending toxins (CDT) are considered prototypic inhibitory cyclomodulins (1, 2). Genes encoding CDTs are widely disseminated among Gram-negative pathogenic bacteria, including *Escherichia coli*, *Campylobacter* spp., *Aggregatibacter actinomycetemcomitans*, *Haemophilus ducreyi*, *Salmonella enterica* serovar Typhimurium, and *Shigella* spp. (3).

The holotoxin is a heterotrimer of three protein subunits, CdtA, CdtB, and CdtC. These are encoded by three adjacent, sometimes slightly overlapping genes (4). CdtB is the active subunit, possessing DNase activity and sharing homology with the mammalian DNase I (5). There is evidence that in case of *Campylobacter* spp., *A. actinomycetemcomitans*, and *H. ducreyi*, the other two subunits play a role in the transport of CdtB into the target cell (3). Upon entering the eukaryotic cell, CdtB causes DNA damage, which in turn causes cell cycle arrest between the G_2 and M phases (6). Double-stranded DNA damage leads to the distension and subsequent death of the target cell (7). The characteristic distending transformation has been demonstrated on multiple cell lines by several studies (7, 8).

CDT was first identified in *E. coli* (Johnson and Lior) (9), and its production has been associated with several pathotypes, e.g., enterohemorrhagic (EHEC) and enteropathogenic (EPEC) (3). So far, five types have been associated with *E. coli*, termed CDT-I to CDT-V (4, 6, 10, 11, 12). The genomic localization of *cdt* alleles and their association with mobile genetic elements was reported by several groups. Accordingly, Pérès et al. localized *cdt-III* to a large conjugative virulence plasmid (6). *cdt-I* and *cdt-IV* are encoded by lambdoid prophages (13, 14), while the *cdt-V* operon is flanked by P2-like phage sequences (11, 14, 15, 16), but the corresponding nucleotide sequences deposited show only small portions of the phage genes *repA* (replication gene A) and *Q* (a capsid gene).

The presence of CDT-V in Shiga-toxigenic *E. coli* (STEC) strains of various serotypes, both of clinical (17, 18) and nonclinical (19) origin, has been reported. CDT-V has also been associated with strains of other serotypes and pathotypes associated with human diarrhea (16, 20), where it is the only known cytotoxin of the respective strains, underlining the importance of CDT-V as a virulence factor. Therefore, it is imperative to obtain more information on the potential mobility of the CDT-V-encoding operon in non-STEC pathotypes and determine whether it is phage associated.

The aim of this study was to characterize the P2-like phage sequence context flanking the *cdt-V* operon in strain T22, an *E. coli* O157:H43 strain of atypical pathotype (*eae* and *stx* negative), and to monitor the presence of characteristic regions in other pathogenic and nonpathogenic CDT-V-positive strains and in additional K-12 *E. coli* strains.

MATERIALS AND METHODS

Bacterial strains. Strains used in this study are listed in Table 1. Strains were grown on lysogeny broth (LB) agar plates or bromothymol blue agar plates.

Cosmid clone library construction and screening. Genomic DNA was isolated from strain T22 of *E. coli* O157:H43 with the Sigma genomic DNA kit (Sigma-Aldrich, St. Louis, MO). The preparation of the cosmid clone library was performed with the pWEB-TNC cosmid cloning kit (Epicentre, Madison, WI) according to the manufacturer's instructions, with the modification that genomic DNA was subjected to a partial digestion with restriction endonuclease MboI (Fermentas, Vilnius, Lithuania). All together, 1,000 transformant colonies were identified (cosmid library) and screened for the presence of *cdt-V* using the primers indicated in Table 2.

Received 8 April 2013 Accepted 6 June 2013
Published ahead of print 14 June 2013
Address correspondence to István Tóth, tothi@vmri.hu.

doi:10.1128/AEM.01134-13

TABLE 1 Bacterial strains used in this study and the presence of P2-like genes and regions in each of them

Strain	Serotype	Pathotype[c]	CDT type	Phylogenetic group[d]	Presence/absence[b] of:																				Reference
					Spacer *Q*	Spacer	*cdtA* spacer	*cdtc-repA*	*P*	*P-Q*	*repA*	*repA* distant	*C*	*O*	Short ORFs	Baseplate *J*	Tail	Terminase	*D*	*FI*	Capsid	Lysin	*cpxP* integrase	*ogr-fieF*	
T22[a]	O157:H43	Atypical	V	B1	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	23
T16[a]	O157:H43	Atypical	V	B1	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	−	+	+	+	+	23
T50[a]	O157:H43	Atypical	V	B1	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	−	+	+	+	+	23
T34[a]	O157:H9	Atypical	V	B1	−	+	+	+	+	+	+	+	+	+	+	+	+	+	+	−	+	+	+	+	23
T49[a]	O157:H37	Atypical	V	B1	−	+	+	+	+	+	+	+	+	+	+	+	+	+	+	−	+	+	+	+	23
T4	O157:H12	Atypical		A	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	23
B20	O157:H12	Atypical		A	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	23
B47	O157:NM	Atypical		B1	−	−	−	−	+	+	+	+	+	+	+	+	+	+	+	−	−	+	+	+	23
B54	O157:H12	Atypical		A	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	23
E6468/62	O86:H34	EPEC	I	ND	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	4
BM2-10	O88	Diarrheagenic	III	ND	−	−	−	−	−	−	−	+	−	−	−	−	+	−	−	−	−	−	−	−	35
28C	O75:K95	ExPEC	IV	ND	−	−	−	−	+	+	+	+	−	−	+	+	+	+	+	+	−	+	−	−	14
703/88[a]	O157:NM	EHEC	V	D	−	+	+	+	+	+	+	+	−	+	+	+	+	+	+	+	−	+	−	−	14
702/88[a]	O157:NM	EHEC	V	D	−	+	+	+	+	+	+	+	−	+	+	+	+	+	+	+	−	+	−	−	14
493/89[a]	O157:NM	EHEC	V	D	−	+	+	+	+	+	+	+	−	+	+	+	+	+	+	+	+	+	−	−	36
E2348/69	O127:H6	EPEC		B2	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	37
Sakai	O157:H7	EHEC		D	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	38
C600	K-12	Commensal		ND	−	−	−	−	−	−	−	+	−	−	−	−	−	−	−	−	−	−	−	−	39
ER2738	K-12	Commensal		ND	−	−	−	−	−	−	−	+	−	−	−	−	−	−	+	−	−	−	−	−	40

[a] Strain was used in the phage induction experiments.

[b] "Spacer" refers to the region between *cdtA* and capsid packaging protein gene *Q*, which has no associated function so far. "Short ORFs" refers to the region spanning 4 ORFs from a replication gene (*repB*) to a hypothetical zinc finger protein. +, presence; −, absence.

[c] EPEC, enteropathogenic *Escherichia coli*; EHEC, enterohemorrhagic *Escherichia coli*; ExPEC, extraintestinal pathogenic *Escherichia coli*.

[d] ND, not determined.

TABLE 2 Primers used for the investigation of P2-like regions

Primer name	Target gene(s)	Sequence 5′ to 3′	Position (bp)[a]	Reference
FI_fw	Tail sheath monomer FI gene	TGCGTGTGGAAGACGGCACC	24612–24631	This study
FI_rev		CGGCCTTGAGGGTTTCCGCAT	25413–25433	This study
Latecontrol_D_fw	Late control protein D gene	TTCGTGGCGCTGTCCTGACG	29694–29713	This study
Latecontrol_D_rev		TGGAGAACTCCGCAACGCCC	30448–30467	This study
Terminase_fw	Terminase gene	GGGAGCCGAACGGATTGGCG	10561–10580	This study
Terminase_rev		CTTGAGCACCGCATCCGCGA	11298–11317	This study
Capsid_fw	Capsid gene	CCGTCGGTCACCCAGACCCT	12975–12994	This study
Capsid_rev		AGCGGCCAGCATTTCGCTGT	13655–13636	This study
Lysin_fw	Lysin gene	ATGCTGGCCGTGTCCGAAGG	15834–15853	This study
Lysin_rev		CTGACCGTAACCGGCACCCG	16204–16223	This study
Tail_protein_fw	Phage tail gene	AGGGCGCTGACTGATGCCGT	17271–17290	This study
Tail_protein_rev		AGCTCCATCGGGCGGGTGAC	17655–17674	This study
Baseplate_J_fw	Baseplate gene *J*	GGTTGCCCGTACCCTGACGC	19335–19354	This study
Baseplate_J_rev		GACAGACGGATGTCGCGCCC	19960–19979	This study
L413C_specific_fw	*C* (repressor)	AGGATCCAGCCCTTTCTAAA	1632–1651	This study
L413C_specific_rev		AATCCTTTTCGCGGAGTGG	1243–1261	This study
gpO_fw	Capsid scaffolding protein O gene	TTTCGTATCGGCGTTGAGG	11984–12002	This study
gpO_rev		TGTTTTTCAGGCGGGTGAA	12710–12728	This study
cpxP2_fw	*cpxP* integrase gene	AGCCATATGTTCGACGGCAT	4327924–4327943*	This study
integrase_rev		AGGCCAGCGCCAAATTATTC	439–458	This study
P2_cdta_up_novel_fw	Spacer region gene	CCAAGAGCACGCCAGCACTGA	8727–8747	This study
P2_cdta_up_novel_rev		TGGCAAATCTCCTTTGGGCTAGTG	8261–8284	This study
Cdta_up_novel_overlap_fw	*cdtA* spacer gene	TCAATCGCGTTTTGCACTCACGG	8500–8522	This study
Cdta_up_novel_overlap_rev		AGGTCCAGCCCCGGGTAATGG	8018–8038	This study
P2_novel_overlap_fw	Capsid packaging protein Q spacer gene	ACAAGGGAAAGCCGACGGCAT	9259–9279	This study
P2_novel_overlap_rev		TCAGTTATCGTCAGTGCTGGCGTGC	8717–8741	This study
Cdtc_P2_overlap_fw	*cdtC* replication gene *A*	CATACCTTCAACAACAGGTGCGGT	6227–6250	This study
Cdtc_P2_overlap_rev		TACCGGTTCCCACGCCTTCTG	5591–5611	This study
P2_short_orfs_fw	Short ORFs	CGCCGTCAGGTTGGCGCAA	2367–2385	This study
P2_short_orfs_rev		AACCAGCGCAGAAGCACCGC	3281–3300	This study
ogr_fw	*ogr-fieF*	ACGTGAATTGCAGCGCCACG	30840–30859	This study
fieF_rev		CCAGCGCCAGCAGAATTGCG	4361610–4361629*	This study
P2_PQ_fw	Capsid packaging protein Q gene	AACACCGCAACCTGCGGCAA	9974–9993	This study
P2_PQ_rev		CCACCAGTAAACATCCTCT	9579–9597	This study
P2_P_fw	Terminase protein P gene	AACCTGTTCATGTGTGAA	10676–10693	This study
P2_P_rev		AACGAGCTGGTGATGTCGG	10191–10209	This study
P2_repA_fw	Replication gene *A*	TTTCGATAATCTGGTTACG	4915–4933	This study
P2_repA_rev		TATGGTCATGTTGCTGGCGC	4213–4232	This study
P2_repA_ext_fw	Replication gene *A* distant region	TTGCTTCAAAGACTGACTG	3376–3394	This study
P2_repA_ext_rev		GAAATGAACGACGTCGAACGT	2298–2318	This study
CDT-IIIs	*cdt-VB*	GAAAGTAAATGGAATATAAATGTCCG	7291–7316	11
CDT-IIIas		TTTGTGTCGGTGCAGCAGGGAAAA	6761–6784	11

[a] The reference sequence for the primers indicated with an asterisk is CP002967 (whole genome of *Escherichia coli* strain W). Positions are given in reference to GenBank accession number KC618326.1, except where indicated otherwise.

Sequencing. Cosmid DNA was isolated by using the alkaline lysis method (21), genomic DNA was isolated by using the GenElute bacterial genomic DNA kit (Sigma-Aldrich, St. Louis, MO), and both were sequenced at the Biological Research Center (Szeged, Hungary) by using the Ion Torrent Personal Genome Machine (PGM) next-generation sequencer as well as traditional Sanger-based capillary sequencing. The average coverage for the prophage region was 112×. Trimming and assembly were performed manually and by using CLC Genomics Workbench version 6.0.1. Nucleotide sequence analysis and searches for open reading frames (ORFs) and homologous DNA sequences in the EMBL and GenBank database libraries were performed with the tools available from the National Center for Biotechnology Information (www.ncbi.nlm.nih.gov) together with Vector NTI and CLC Bio Genomics Workbench softwares.

PCR screening for flanking regions. *E. coli* strains representing different serotypes and pathotypes (Table 1) were tested by PCR for the presence of characteristic P2-like phage genes. Primers were designed with the aid of PrimerBLAST, available from the National Center for Biotechnology Information. The general PCR conditions were an initial denaturation of 3 min at 94°C and then 30 cycles of denaturation for 30 s at 94°C, followed by annealing for 30 s at 59 to 60°C (depending on the primer pair) and extension for 60 s at 72°C. The final extension time was 5 min. The primers used in the reactions are listed in Table 2.

Phage induction experiments. Phage induction was carried out for the 8 strains marked in Table 1, using either mitomycin C (0.5 μg ml^{-1}) or norfloxacin (1.25 μg ml^{-1}) as inducing agents. Induction was also attempted by UV irradiation as described by Hertman and Luria (22), with the modification that doses were 15, 10, and 5 s long with a 30-W lamp. *E. coli* K-12 strains ER2738, C600, and *Shigella sonnei* strain 866 (19), kindly provided by Maite Muniesa, were used as indicator strains. As a positive control for phage induction, the *E. coli* O157:H7 strain 34, harboring an inducible Stx2 phage (23), was used.

Nucleotide sequence accession number. The nucleotide sequence of the P2-like prophage has been deposited into GenBank (KC618326.1).

RESULTS

Characteristics of the P2-like prophage carrying the *cdt-V* cluster. In order to determine the sequence of the *cdt-V* operon and its exact position in the genome of *E. coli* O157:H43 strain T22, a clone of the cosmid library positive for *cdt-V*, as well as the corresponding region of the genomic DNA, was sequenced, annotated, and deposited into GenBank. The length of the prophage containing the *cdt-V* operon is 31.2 kb. The *cdt-V* operon of strain T22 shows the highest nucleic acid similarity to the *cdt* operon of strain AH-10 (GenBank no. AB472839). The sequence of *cdtB*, coding for the active toxin subunit, is highly homologous to the AH-10-specific sequence, with only 4 synonymous single-nucleotide polymorphisms (SNPs). The CDT-V-encoding operon in strain T22 is flanked by P2-like prophage sequences. The majority of P2-like genes carried by T22 have the highest similarity to the bacteriophage L-413C (97% similarity on average at the nucleotide level; GenBank no. AY251033). The GC content of the *cdt-V* operon is 41%, whereas the prophage genes exhibit an average GC content of 53%. The structural organization of the P2-like phage genes is reminiscent of the P2-like prophage in *E. coli* strain W (GenBank no. CP002967). DNA sequence comparison between the P2-like sequences of strains T22 and W revealed that there is a high degree of nucleotide homology (94 to 99% identity) between their genes (Table 3). The tail fiber-encoding genes (ORFs 33 to 36) located in the region ranging from bp positions 20720 to 23571 are different from all published P2-like prophages. The respective phage tail fiber genes of the prophage in the genome of *E. coli* strain UMNK88 (accession no. CP002729; with the homologue of ORF 34 being present in two copies) are their closest homologues, with a nucleotide identity of 92 to 99%. The region between bp positions 22000 and 22192 is entirely missing in all other P2-like prophage sequences available in GenBank, whereas it is present in the prophage of UMNK88. The ORFs predicted in the P2-like prophage of *E. coli* O157 strain T22 are listed in Table 3.

PCR scanning of P2-like genes. The dissemination of characteristic *cdt-V*-flanking genes ($n = 20$) in a collection of CDT-producing and nonproducing strains was examined by PCR. At least one gene from each P2-like functional gene cluster was investigated. The list of results, together with the genotype of the strains, is given in Table 1. Fifteen out of 20 primer pairs designed specifically for the flanking regions (Table 2) yielded PCR products in all CDT-V-positive strains. The overlapping primer pairs covering the Q capsid gene and the spacer region between the Q gene and *cdtA* yielded only a PCR product in the O157:H43 strains tested, and the tail sheath monomer-encoding FI gene could be amplified only from strain T22 and the O157:NM CDT-V strains. Strains carrying different *cdt* types and the CDT-negative strains carried fewer P2-like genes. Interestingly, the CDT-negative atypical strain B47 and the CDT-IV-positive strain 28C were positive for 15 and 12 of the investigated sequences, respectively. Sequencing of the amplicons produced by the P2_PQ primer pair (Table 2) showed that in this 376-bp region of the Q capsid gene, there was a maximum of 14 SNPs, with only 3 of them leading to amino acid changes.

Phage induction experiments from CDT-V-positive strains. In order to induce lysogenic bacteriophages and to isolate corresponding phage particles from the CDT-V-positive strains, phage induction experiments were carried out with 8 CDT-V-positive strains, marked in Table 1, by using UV light as well as mitomycin C and norfloxacin as inducing agents. Among CDT-V-positive strains, only EHEC O157:NM strains 702/88 and 493/89 released phages upon induction with mitomycin C, which were able to infect and lyse *E. coli* strain C600. PCR investigations revealed that none of these phages carried *cdt-V* genes (data not shown). These data indicate that the P2-like prophages harboring the *cdt-V* operon cannot be induced from these *stx*-negative and *cdt-V*-positive *E. coli* O157 strains.

DISCUSSION

We sequenced for the first time a whole P2-like prophage from the genome of *E. coli* O157:H43 strain T22 containing a *cdt-V* operon. The *cdt-V* sequence in strain T22 is highly similar to that of the *cdt-V* operon in strain AH-10 (16). Our results confirm and expand our knowledge based on previous reports that the *cdt-V* operon is flanked by P2-like phage sequences (12, 16) and on a more recent finding of an inducible P2-like phage carrying the *cdt-V* operon (19). The proximal P2-like regions in *E. coli* T22 are highly similar to the sequences published by the above-mentioned authors; however, in our study, the entire prophage genome was resolved and annotated. Among the lytic P2-like bacteriophages, phage L-413C (induced originally from a *Yersinia pestis* strain [24, 25]) has the highest homology compared to the prophage sequence of strain T22 (GenBank no. AY251033). Among other available prophage sequences, the P2-like prophage of *E. coli* O157:H7 strain TW14359 (GenBank no. CP001368.1) is the closest homologue. Compared to all the P2-like prophage sequences available in GenBank, the level of homology between the individual structure genes and their closest homologues is between 94 and 100% (Table 3). The Z/fun region in the original P2 phage, situated between the tail fiber gene *G* and the tail sheath gene *FI*, is known to host insertions of foreign genes (26). In the case of strain T22, the P2-like prophage does not contain a gene homologous to the G gene. Instead, the tail fiber genes between bp positions 20720 and 23571 (ORFs 33 to 36; see Table 3) seem to have a different origin in this strain, as they show only partial homology to P2-like phages or prophages, their closest homologues being tail fiber genes of a prophage carried by *E. coli* strain UMNK88. Among the flanking ORFs, ORF 32 corresponds to gene *H* (upstream of *G*) and ORF 37 to gene *FI*; therefore, the insertion of ORFs 33 to 36 occurred essentially in the Z/fun region, with the additional deletion of gene *G*. Nilsson and coworkers suggested site-specific recombination as a possible mechanism for the insertion of foreign genes in this region (27). The potential recombination of prophages in *E. coli* O157 was already reported (28), and it was also suggested that the mosaic structure of prophages is the result of extensive exchange of genetic material among different bacteriophages and also their hosts (13), like in the case of lambdoid prophages harboring the *cdt-I* and *cdt-IV* operons (13, 14). The genome sequencing project of strain T22 (29) confirmed that the integration site of the prophage is between the *cpxP* and *fieF* genes. This is the same site where P2-like prophages can be found in a further four *E. coli* strains according to publicly available GenBank entries (accession no. CP002967.1, CP002797.2, CP001969.1, and CP000970.1).

The integration site of the *cdt-V* operon within the prophage, the TO region, is known to be an integration hot spot for foreign genes in P2-like phages (26). This region is named after the phage immunity genes *tin* and *old*, encoded by the lytic phages in this region. It has been suggested that this site is a potential carrier of

TABLE 3 Functional annotation of the P2-like prophage in *E. coli* O157:H43 strain T22[a]

ORF	Feature or product of sequence	Position		Size (bp)	No. of identical nt/total no. of nt, % identity	Reference accession no.	Strain
		Start	End				
1	Integrase	187	1167	981	980/981, 99	CP002797.2	NA114
2	Immunity repressor	1237	1530	294	**294/294, 100**	AY251033.1	Bacteriophage L-413C
3	Repressor, prophage excision	1667	1939	273	**273/273, 100**	CP002797.2	NA114
4	Unknown	1915	2112	198	196/198, 99	CP002797.2	NA114
5	DNA replication	2109	2609	501	**501/501, 100**	CP001368.2	TW14359
6	Unknown	2673	2897	225	**223/225, 100**	CP000800.1	E24377A
7	Unknown	2897	3199	303	290/303, 96	CP002967.1	W
8	Hypothetical zinc finger protein	3199	3423	225	221/225, 98	CP002185	W
9	Unknown	3420	3695	276	274/276, 99	CP002967.1	W
10	DNA replication	3685	5982	2,298	2191/2262, 97	CP002185	W
11	Cytolethal distending toxin C subunit	6058	6603	546	545/546, 99	AB472839	AH-10
12	Cytolethal distending toxin B subunit	6618	7427	810	809/810, 99	AB472860	AH-16
13	Cytolethal distending toxin A subunit	7424	8200	777	**777/777, 100**	JF461073	Bacteriophage fi125
14	Hypothetical protein	8779	8940	162	161/162, 99	AB472870	AH-26
15	Phage-related capsid packaging protein	8979	10013	1,035	1009/1035, 97	CP000970	SMS-3-5
16	Phage terminase, ATPase subunit	10014	11786	1,773	1724/1774, 97	AY251033.1	Bacteriophage L-413C
17	Phage capsid scaffolding protein	11894	12814	921	896/921, 97	CP002185	W
18	Phage major capsid protein	12873	13946	1,074	1060/1074, 97	CP002185	W
19	Phage terminase, endonuclease subunit	13951	14693	744	727/744, 98	AJ298566.1	Bacteriophage 299
20	Phage head completion, stabilization	14793	15302	510	499/510, 98	CP000970	SMS-3-5
21	Phage tail completion protein	15302	15505	204	202/204, 99	AY251033.1	Bacteriophage L-413C
22	Phage holin	15544	15789	246	**246/246, 100**	AY251033.1	Bacteriophage L-413C
23	Phage lysin	15789	16286	498	492/498, 99	CP002967.1	W
24	Phage holin, endolysin	16301	16726	414	420/426, 99	CU928161	S88
25	Phage spanin, Rz	16714	17139	426	412/426, 97	CP000970	SMS-3-5
26	Phage outer membrane lipoprotein	17126	17284	159	157/159, 99	CP002967.1	W
27	Phage tail protein	17247	17714	468	461/468, 99	CP002967.1	W
28	Phage tail completion protein	17707	18159	453	441/453, 97	CP002967	W
29	Baseplate assembly	18226	18861	636	622/636, 98	CP002967.1	W
30	Baseplate assembly	18858	19205	348	342/348, 98	CU928162	ED1a
31	Baseplate assembly	19210	20118	909	890/909, 98	CP001969	IHE3034
32	Phage tail fiber	20111	20722	612	603/612, 99	CP001368	TW14359
33	Phage tail fiber	20719	22218	1,500	1437/1502, 96	CP002729	UMNK88
34	Putative tail fiber assembly protein	22218	22820	603	595/603, 99	CP002729	UMNK88
35	Tail fiber assembly protein	22792	23235	444	433/444, 98	CP001368	TW14359
36	Hypothetical protein	23256	23570	315	289/315, 92	CP002729	UMNK88
37	Phage DNA invertase	23713	24207	495	468/499, 94	CP002967	W
38	Phage tail sheath protein	24362	25552	1,191	1173/1191, 98	HE616528	*Shigella sonnei* 53G
39	Phage tail tube protein	25565	26083	519	518/519, 99	CU928161	S88
40	Tail protein	26140	26415	276	267/276, 97	CP002516	KO11
41	Putative phage tail protein	26412	26567	156	**156/156, 100**	CP004009	APEC O78
42	Phage protein	26560	29007	2,448	2390/2448, 98	CU928162	ED1a
43	gpU (tail protein)	29022	29501	480	471/480, 98	CP002967	W
44	Gene *D* protein	29501	30664	1,164	1137/1164, 98	CU928161	S88
45	Putative positive regulator of lysis	30671	30964	294	288/294, 98	CP000970	SMS-3-5

[a] Positions refer to GenBank accession number KC618326.1. Homologies of 100% are in bold.

advantageous genes for the host (26). The carriage of a potent virulence-associated gene, like *cdt-V*, provides evidence for this notion. It also has to be mentioned that, so far, CDT-V is the only established virulence factor with its genes inserted into the TO region of P2-like prophages. In complete P2-like prophage sequences available in GenBank, the TO region is usually occupied by a gene coding for a hypothetical protein, or in the case of EHEC O157:H7 strain TW14359, the genes of a putative virulence factor, one of which shows partial homology to a eukaryotic serine esterase and a gene of *Bacillus amyloliquefaciens* (30). Genes encoding reverse transcriptases have been also reported in the TO regions of P2-like prophages found in the genomes of ECOR strains (31). The foreign origin of the *cdt-V* operon is also supported by the considerable difference between the GC content of the operon and the prophage (41% versus 53%).

The fact that no phages were detected from any of the induced

cdt-V-positive strains, neither by induction with UV light, mitomycin C, or norfloxacin, is in harmony with current knowledge that P2-like phages are classified as noninducible in *E. coli* (26). This finding also suggests that the inducible prophages found in STEC strains (19) may represent an earlier evolutionary stage of the *cdt-V*-carrying P2-like prophages in *E. coli* and that the phages, which we were unable to induce, could have become temperated by adapting to the host organism.

The P2-like prophage sequences seem to be characteristic for the CDT-V-positive strains, with few differences that can be attributed to the adaptation processes in the various hosts. The presence of the L-413C-like C gene in the *stx*- and *eae*-negative O157 CDT-V strains, relative to EHEC CDT-V strains, can be explained with the high variability of this gene, which therefore can be used as a marker to identify and distinguish different P2-like phage variants (26). Our results suggest that while the *stx*-negative *cdt-V* O157:H43 strains uniformly contain a P2-like variant with an L-413C-like C gene, the P2-like prophages in the other strains may carry different variants of this gene, as they represent other evolutionary lineages of P2-like phages. It will be an important future task to evaluate whether different P2-like bacteriophage variants may be associated with individual *E. coli* clonal groups and/or with different virulence factors in their TO region. An interesting finding was the carriage of P2-like genes by strain 28C, an extraintestinal pathogenic *E. coli* strain, which harbors the *cdt-IV* operon flanked by lambda phage-like genes (14). This observation may indicate that the P2-like genes could be located elsewhere in the *E. coli* 28C genome. Alternatively, recombination between a P2-like and a lambdoid phage cannot be excluded. Another example of P2-like phages serving as vehicles for foreign DNA can apparently be found in the CDT-negative atypical O157:NM strain B47, which also carries most of the P2-like prophage genes investigated and is where the integration site between *cpxP* and *fieF* also seems to be occupied. Strain B47 probably also carries a nearly complete P2-like prophage similar to other investigated O157 CDT-V strains, with as-yet-unknown genes inserted into its integration hot spots. On the other hand, the PCRs specific for the integration site of the P2-like prophage were negative in the case of the O157:NM EHEC strains, suggesting a different integration site for the prophage in these isolates. The atypical O157 strains harboring P2-like genes are phylogenetically distinct from those that do not carry them. While the CDT-negative O157 strains (T4, B20, and B54; all representing the O157:H12 serotype) belong to phylogenetic group A, the CDT-positive O157 strains (T22, T16, T34, T49, T50), as well as strain B47, belong to phylogenetic group B1 (Table 1) (32).

It has been proposed that the presence of P2-like phage sequences suggests the common acquisition of the *cdt-V* operon in O157:NM EHEC strains after the lineage has diverged from the O157:H7 strains (12). While there is indeed a strong association between O157:NM strains and the presence of CDT-V (17, 19, 33), several strains from other serotypes, both from healthy cattle (23, 33) and from cases of human diarrhea (15, 16, 17, 34), express CDT-V. These findings, together with the results of our study, indicate that while the *cdt-V* genes are rather conserved, the carrier P2-like phages became diverse during the evolution of their hosts, and this event in most cases may have resulted in loss of their mobility. These facts suggest a more diverse evolutionary and/or transductional history of the *cdt-V* operon and its carrying of P2-like prophages. The highly conserved *cdt-V* operon within more variable and potentially inactivated bacteriophage genomes may result from selective pressure to maintain a functional *cdt* gene cluster and to stabilize this cargo determinant by inactivation of this bacteriophage genome. Further investigations of flanking regions and P2-like prophage sequences in CDT-V-positive strains are expected to help in clarifying the evolutionary background of the distribution of these variants.

ACKNOWLEDGMENT

This work was supported by the Hungarian Research Foundation (OTKA), grant number K 81252. U.D. was supported by the German Research Foundation (DO 789/4-1; SFB1009 TP B05).

REFERENCES

1. **Nougayrède J, Taieb F, De Rycke J, Oswald E.** 2005. Cyclomodulins: bacterial effectors that modulate the eukaryotic cell cycle. Trends Microbiol. **13:**103–110.
2. **Oswald E, Nougayrède J, Taieb F, Sugai M.** 2005. Bacterial toxins that modulate host cell-cycle progression. Curr. Opin. Microbiol. **8:**83–91.
3. **Jinadasa RN, Bloom SE, Weiss RS, Duhamel GE.** 2011. Cytolethal distending toxin: a conserved bacterial genotoxin that blocks cell cycle progression, leading to apoptosis of a broad range of mammalian cell lineages. Microbiology **157:**1851–1875.
4. **Scott DA, Kaper JB.** 1994. Cloning and sequencing of the genes encoding *Escherichia coli* cytolethal distending toxin. Infect. Immun. **62:**244–251.
5. **Nesić D, Hsu Y, Stebbins CE.** 2004. Assembly and function of a bacterial genotoxin. Nature **429:**429–433.
6. **Pérès SY, Marchès O, Daigle F, Nougayrède JP, Herault F, Tasca C, De Rycke J, Oswald E.** 1997. A new cytolethal distending toxin (CDT) from *Escherichia coli* producing CNF2 blocks HeLa cell division in G_2/M phase. Mol. Microbiol. **24:**1095–1107.
7. **Comayras C, Tasca C, Pérès SY, Ducommun B, Oswald E, De Rycke J.** 1997. *Escherichia coli* cytolethal distending toxin blocks the HeLa cell cycle at the G_2/M transition by preventing cdc2 protein kinase dephosphorylation and activation. Infect. Immun. **65:**5088–5095.
8. **Bielaszewska M, Sinha B, Kuczius T, Karch H.** 2005. Cytolethal distending toxin from Shiga toxin-producing *Escherichia coli* O157 causes irreversible G_2/M arrest, inhibition of proliferation, and death of human endothelial cells. Infect. Immun. **73:**552–562.
9. **Johnson WM, Lior H.** 1988. A new heat-labile cytolethal distending toxin (CLDT) produced by *Escherichia coli* isolates from clinical material. Microb. Pathog. **4:**103–113.
10. **Pickett CL, Cottle DL, Pesci EC, Bikah G.** 1994. Cloning, sequencing, and expression of the *Escherichia coli* cytolethal distending toxin genes. Infect. Immun. **62:**1046–1051.
11. **Tóth I, Hérault F, Beutin L, Oswald E.** 2003. Production of cytolethal distending toxins by pathogenic *Escherichia coli* strains isolated from human and animal sources: establishment of the existence of a new cdt variant (type IV). J. Clin. Microbiol. **41:**4285–4291.
12. **Janka A, Bielaszewska M, Dobrindt U, Greune L, Schmidt MA, Karch H.** 2003. Cytolethal distending toxin gene cluster in enterohemorrhagic *Escherichia coli* O157:H− and O157:H7: characterization and evolutionary considerations. Infect. Immun. **71:**3634–3638.
13. **Asakura M, Hinenoya A, Alam MS, Shima K, Zahid SH, Shi L, Sugimoto N, Ghosh AN, Ramamurthy T, Faruque SM, Nair GB, Yamasaki S.** 2007. An inducible lambdoid prophage encoding cytolethal distending toxin (Cdt-I) and a type III effector protein in enteropathogenic *Escherichia coli*. Proc. Natl. Acad. Sci. U. S. A. **104:**14483–14488.
14. **Tóth I, Nougayrède JP, Dobrindt U, Ledger TN, Boury M, Morabito S, Fujiwara T, Sugai M, Hacker J, Oswald E.** 2009. Cytolethal distending toxin type I and type IV genes are framed with lambdoid prophage genes in extraintestinal pathogenic *Escherichia coli*. Infect. Immun. **77:**492–500.
15. **Bielaszewska M, Fell M, Greune L, Prager R, Fruth A, Tschäpe H, Schmidt MA, Karch H.** 2004. Characterization of cytolethal distending toxin genes and expression in Shiga toxin-producing *Escherichia coli* strains of non-O157 serogroups. Infect. Immun. **72:**181–186.
16. **Hinenoya A, Naigita A, Ninomiya K, Asakura M, Shima K, Seto K, Tsukamoto T, Ramamurthy T, Faruque SM, Yamasaki S.** 2009. Prevalence and characteristics of cytolethal distending toxin-producing *Escherichia coli* from children with diarrhea in Japan. Microbiol. Immunol. **53:**206–215.

17. **Friedrich AW, Lu S, Bielaszewska M, Prager R, Bruns P, Xu J, Tschäpe H, Karch H.** 2006. Cytolethal distending toxin in *Escherichia coli* O157:H7: spectrum of conservation, structure, and endothelial toxicity. J. Clin. Microbiol. **44:**1844–1846.
18. **Osek J.** 2005. Detection of cytolethal distending toxin genes in Shiga toxin-producing *Escherichia coli* isolated from different sources. Bull. Vet. Inst. Pulawy **49:**153–156.
19. **Allué-Guardia A, García-Aljaro C, Muniesa M.** 2011. Bacteriophage encoding cytolethal distending toxin type V gene induced from nonclinical *Escherichia coli* isolates. Infect. Immun. **79:**3262–3272.
20. **Kim J, Kim J, Choo Y, Jang H, Choi Y, Chung J, Cho S, Park M, Lee B.** 2009. Detection of cytolethal distending toxin and other virulence characteristics of enteropathogenic *Escherichia coli* isolates from diarrheal patients in Republic of Korea. J. Microbiol. Biotechnol. **19:**525–529.
21. **Birnboim HC, Doly J.** 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res. **7:**1513–1523.
22. **Hertman I, Luria SE.** 1967. Transduction studies on the role of a rec^+ gene in the ultraviolet induction of prophage lambda. J. Mol. Biol. **23:**117–133.
23. **Tóth I, Schmidt H, Kardos G, Lancz Z, Creuzburg K, Damjanova I, Pászti J, Beutin L, Nagy B.** 2009. Virulence genes and molecular typing of different groups of *Escherichia coli* O157 strains in cattle. Appl. Environ. Microbiol. **75:**6282–6291.
24. **Larina VS, Anisimov PI, Adamov AK.** 1970. A novel strain of plague bacteriophage for identification of *Pasteurella pestis*. Probl. Particularly Dangerous Infect. **11:**132–136.
25. **Garcia E, Chain P, Elliott JM, Bobrov AG, Motin VL, Kirillina O, Lao V, Calendar R, Filippov AA.** 2008. Molecular characterization of L-413C, a P2-related plague diagnostic bacteriophage. Virology **372:**85–96.
26. **Nilsson AS, Haggård-Ljungquist E.** 2007. Evolution of P2-like phages and their impact on bacterial evolution. Res. Microbiol. **158:**311–317.
27. **Nilsson AS, Karlsson JL, Haggård-Ljungquist E.** 2004. Site-specific recombination links the evolution of P2-like coliphages and pathogenic bacteria. Mol. Biol. Evol. **21:**1–13.
28. **Asadulghani M, Ogura Y, Ooka T, Itoh T, Sawaguchi A, Iguchi A, Nakayama K, Hayashi T.** 2009. The defective prophage pool of *Escherichia coli* O157: prophage-prophage interactions potentiate horizontal transfer of virulence determinants. PLoS Pathog. **5:**e10000408. doi:10.1371/journal.ppat.1000408.
29. **Sváb D, Horváth B, Szűcs A, Maróti G, Tóth I.** 2013. Draft genome sequence of an *Escherichia coli* O157:H43 strain isolated from cattle. Genome Announc. **1:**e00263–13. doi:10.1128/genomeA.00263-13.
30. **Kulasekara BR, Jacobs M, Zhou Y, Wu Z, Sims E, Saenphimmachak C, Rohmer L, Ritchie JM, Radey M, McKevitt M, Freeman TL, Hayden H, Haugen E, Gillett W, Fong C, Chang J, Beskhlebnaya V, Waldor MK, Samadpour M, Whittam TS, Kaul R, Brittnacher M, Miller SI.** 2009. Analysis of the genome of the *Escherichia coli* O157:H7 2006 spinach-associated outbreak isolate indicates candidate genes that may enhance virulence. Infect. Immun. **77:**3713–3721.
31. **Odegrip R, Nilsson AS, Haggård-Ljungquist E.** 2006. Identification of a gene encoding a functional reverse transcriptase within a highly variable locus in the P2-like coliphages. J. Bacteriol. **188:**1643–1647.
32. **Sváb D, Tóth I.** 2012. Allelic types of long polar fimbriae in bovine and human *Escherichia coli*. O157 strains. Acta Vet. Hung. **60:**1–15.
33. **Orth D, Grif K, Dierich MP, Würzner R.** 2006. Cytolethal distending toxins in Shiga toxin-producing *Escherichia coli*: alleles, serotype distribution and biological effects. J. Med. Microbiol. **55:**1487–1492.
34. **Bielaszewska M, Stoewe F, Fruth A, Zhang W, Prager R, Brockmeyer J, Mellmann A, Karch H, Friedrich AW.** 2009. Shiga toxin, cytolethal distending toxin, and hemolysin repertoires in clinical *Escherichia coli* O91 isolates. J. Clin. Microbiol. **47:**2061–2066.
35. **De Rycke J, Plassiart G.** 1990. Toxic effects for lambs of cytotoxic necrotising factor from *Escherichia coli*. Res. Vet. Sci. **49:**349–354.
36. **Karch H, Böhm H, Schmidt H, Gunzer F, Aleksic S, Heesemann J.** 1993. Clonal structure and pathogenicity of Shiga-like toxin-producing, sorbitol-fermenting *Escherichia coli* O157:H−. J. Clin. Microbiol. **31:**1200–1205.
37. **Iguchi A, Thomson NR, Ogura Y, Saunders D, Ooka T, Henderson IR, Harris D, Asadulghani M, Kurokawa K, Dean P, Kenny B, Quail MA, Thurston S, Dougan G, Hayashi T, Parkhill J, Frankel G.** 2009. Complete genome sequence and comparative genome analysis of enteropathogenic *Escherichia coli* O127:H6 strain E2348/69. J. Bacteriol. **191:**347–354.
38. **Hayashi T, Makino K, Ohnishi M, Kurokawa K, Ishii K, Yokoyama K, Han CG, Ohtsubo E, Nakayama K, Murata T, Tanaka M, Tobe T, Iida T, Takami H, Honda T, Sasakawa C, Ogasawara N, Yasunaga T, Kuhara S, Shiba T, Hattori M, Shinagawa H.** 2001. Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genomic comparison with a laboratory strain K-12. DNA Res. **8:**11–22.
39. **Appleyard RK.** 1954. Segregation of new lysogenic types during growth of a doubly lysogenic strain derived from *Escherichia coli* K12. Genetics **39:**440–452.
40. **Woodcock DM, Crowther PJ, Doherty J, Jefferson S, DeCruz E, Noyer-Weidner M, Smith SS, Michael MZ, Graham MW.** 1989. Quantitative evaluation of *Escherichia coli* host strains for tolerance to cytosine methylation in plasmid and phage recombinants. Nucleic Acids Res. **17:**3469–3478.