Table 4. Characteristics of the ORFs of *cdtI*-phage genome and results of BLAST search showing reported homology of the ORFs with that of other phages or bacteria.

ORF no.	Gene coordinates and orientation	Gene product Size (aa)	GC%	Possible function	Related phage and bacterial proteins			Source
					Products (size and origin)	GenBank accession no.	BLAST e-value (identity) %	
1	185→724	179	51.11	Small terminase subunit	Hypothetical protein EcolB_01000955 (179 aa; *E. coli* B171 [EPEC O111:NM])	ZP 00712009	4.80E-97 (179/179) 100%	DS[a]
2	733→2832	699	54.43	Large terminase subunit (GpA)	Bacteriophage tail assembly protein (677 aa; *E. coli* B171 [EPEC O111:NM])	ZP 00712010	0 (675/677) 99.7%	DS
3	2829→3041	70	54.79	Hypothetical protein	Hypothetical protein EcolB_01004539 (70 aa; *E. coli* B171 [EPEC O111:NM])	ZP 00708566	3.30E-33 (70/70) 100.0%	DS
4	3041→4549	502	54.67	Portal protein	Bacteriophage capsid protein (526 aa; *E. coli* UTI89 [uropathogenic strains])	YP 544050	0 (502/502) 100.0%	1
5	4563→6521	652	51.56	Protease/scaffold protein	Putative protease/scaffold protein (707 aa; *E. coli* UTI89 [uropathogenic strains])	YP 544051	0 (650/652) 99.7%	1
6	6608→6931	107	51.23	Unknown	Hypothetical protein ECs0830 (122 aa; *E. coli* str. Sakai [O157:H7])	NP 308857	6.40E-53 (106/107) 99.1%	2, 3
7	6924→7184	86	54.41	Unknown	ABC-type multidrug transport system (91 aa; *E. coli* B171 [EPEC O111:NM])	ZP 00712015	7.60E-38 (79/81) 97.5%	DS
8	7210→7788	192	55.44	DNA packaging, phage assembly	Hypothetical protein (192 aa; *E. coli* B171 [EPEC O111:NM])	ZP 00712016	2.40E-103 (192/192) 100%	DS
9	7773→8186	137	53.38	DNA packaging, phage assembly	Hypothetical protein EcolE1_01000052 (137 aa; *E. coli* E110019 [atypical EPEC O111:H9])	ZP 00723766	1.00E-74 (136/137) 99.3%	DS
10	8197→8940	247	55.24	Cell motility and secretion	Bacterial surface proteins containing Ig-like domains (247 aa; *E. coli* B171 [EPEC O111:NM])	ZP 00712018	9.70E-137 (245/247) 99.2%	DS
11	8998→9387	129	52.82	Structural component	Putative tail component of prophage CP-933K (129 aa; *E. coli* UTI89 [uropathogenic strains])	YP 544057	2.90E-66 (127/129) 98.4%	1
12	9396→9725	109	56.06	Minor tail protein	Putative minor tail protein (Bacteriophage lambda GpT) (109 aa; *E. coli* str. Sakai [O157:H7])	NP 308863	1.90E-57 (109/109) 100%	2, 3
13	9697→12753	1018	53.42	Tail protein	Putative tail length tape measure protein precursor (1018 aa; *E. coli* UTI89 [uropathogenic strains])	YP 544059	0 (1013/1018) 99.5%	1

14	12753→13082	109	52.24	Tail tip assembly	putative minor tail protein (Bacteriophage lambda GpM) (109 aa; *E. coli* str. Sakai [O157:H7])	NP 308865	7.00E-60 (109/109) 100%	2, 3
15	13092→13790	232	55.36	Tail tip assembly	Phage minor tail protein L (232 aa; *E. coli* B7A [ETEC O148:H28])	ZP 00714017	4.70E-135 (232/232) 100%	DS
16	13819→14538	239	59.58	Tail tip assembly	Putative tail fiber component K of prophage (261aa; *E. coli* UTI89 [uropathogenic strains])	YP 544062	7.60E-144 (234/239) 97.9%	1
17	14436→15083	215	59.41	Tail tip assembly	Putative tail component of prophage CP-933K (215 aa; *E. coli* UTI89 [uropathogenic strains])	YP 544063	1.40E-111 (203/215) 94.4%	1
18	15144→18641	1165	55.95	Tail tip assembly, central tail fiber	Putative host specificity protein (1165 aa; *Shigella sonnei* Ss046)	YP 311288	0 (1136/1165) 97.5%	4
19	18711→19310	199	55.5	Lom protein	Outer membrane receptor proteins (Lom protein) (199 aa; *E. coli* B171 [EPEC O111:NM])	ZP 00712028	3.10E-112 (196/199) 98.5%	DS
20	19525→20688	387	59.71	Tail fiber protein	Putative tail fiber protein (437aa; *E. coli* str. Sakai [O157:H7])	NP 310968	0 (370/387) 95.6%	2, 3
21	20651→20959	102	48.54	Unknown	Putative tail fiber protein of bacteriophage BP-933W (102 aa; *E. coli* EDL933 [O157:H7])	NP 286994	6.90E-52 (100/102) 98.0%	5
22	20065←20823	252	59.95	Unknown	Hypothetical protein Stx2Ip026 (284 aa; Stx2 converting bacteriophage I)	NP 612905	1.90E-111 (196/225) 87.1%	6
23	21066→21947	293	39.46	Unknown	Non-LEE-encoded type III effector H (293 aa; *E. coli* str. Sakai [O157:H7])	NP 286534	2.40E-154 (268/293) 91.5%	3, 7
24	22171→22998	275	40.58		Truncated nonfunctional Cif (275 aa; *E. coli* [O26:H11])	AAN07904	2.80E-154 (270/275) 98.2%	8
25	23204←23476	90	45.05	Unknown	Unkown			
26	23437←23616	59	40.00	Unknown	Unkown			
27	24250→24963	237	43.28	CdtA	CdtA (237 aa; *E. coli* [O86:H34])	AAD10621	1.80E-137 (236/237) 99.6%	9
28	24960→25781	273	42.34	CdtB	CdtB (273 aa; *E. coli* [O86:H34])	AAD10622	1.40E-153 (273/273) 100.0%	9
29	25778→26350	190	42.41	CdtC	CdtC (190 aa; *E. coli* [O86:H34])	AAD10623	1.30E-106 (190/190) 100.0%	9
30	26458←26706	82	36.76		Hypothetical protein EcolB_01000978 (82 aa; *E. coli* B171 [EPEC O111:NM])	ZP 00712031	3.60E-40 (82/82) 100.0%	DS

31	26790→26993	67	36.76	Unknown	Unkown			
32	27218→28435	405	45.16	Integrase	Integrase-like protein (P4-like integrases) (404 aa; *E. coli* UTI89 [uropathogenic strains])	YP 544016	0 (399/404) 98.8%	1
33	28569→30701	710	33.99	Unknown	Unknown			DS
34	30809←31003	64	41.03	Unknown	Unknown			DS
35	31003←31623	206	50.56	Unknown	Valyl-tRNA synthetase (206 aa; *E. coli* F11 [ExPEC O6:H31])	ZP 00720184	4.90E-119 (197/206) 95.6%	DS
36	31623←31985	120	45.73	Unknown	Unknown (ofr29) (120aa; *Shigella flexner*i bacteriophage V)	AAL89431	1.70E-66 (116/120) 96.7%	10, 11
37	31976←32512	178	46.18	Unknown	Hypothetical protein UTI89_C5091(yfdR2) (Predicted hydrolases of HD superfamily) (178 aa; *E. coli* UTI89 [uropathogenic strains])	YP 544022	5.00E-99 (175/178) 98.3%	1
38	32640←33464	274	45.45	Unknown	Unknown (orf31) (274 aa; *Shigella flexneri* bacteriophage V)	AAL89433	7.00E-150 (271/274) 98.9%	10, 11
39	33531←33893	120	50.96	Unknown	Hypothetical protein UTI89_C5093 (182 aa; *E. coli* UTI89 [uropathogenic strains])	YP 544024	2.20E-61 (120/120) 100.0%	1
40	33705→34160	151	52.19	Unknown	Hypothetical protein c3199 (227 aa; *E. coli* CFT073 [uropathogenic strains])	AAN81651	8.80E-63 (115/125) 92.0%	12
41	34560←35234	224	44.89	Repressor (cI)	Repressor (SOS-response transcriptional repressors) (224 aa; *Shigella flexneri* bacteriophage V)	AAL89436	3.20E-130 (223/224) 99.6%	10, 11
42	35325→35525	66	42.79	Repressor (cro)	Repressor (cro) (66 aa; *Shigella flexneri* bacteriophage V)	AAL89437	1.40E-31 (66/66) 100%	10, 11
					Hypothetical protein UTI89_C3003 (167 aa; *E. coli* UTI89 [uropathogenic strains])	YP 541990	1.40E-31 (66/66) 100%	1
43	35557→36120	187	51.77	Unknown	Hypothetical protein EcolB_01000938 (187 aa; *E. coli* B171 [EPEC O111:NM])	ZP 00711993	7.60E-104 (187/187) 100.0%	DS
44	36117→37265	382	48.36	Antirepressor	Uncharacterized phage-encoded protein (382aa; *E. coli* B171 [EPEC O111:NM])	ZP 00711994	0 (381/382) 99.7%	DS
45	37262→37486	74	50.22	Unknown	Hypothetical protein UTI89_C5099 (74 aa; *E. coli* UTI89 [uropathogenic strains])	YP 544030	3.40E-38 (74/74) 100.0%	1
46	37483→38367	294	49.49	Unknown	Putative replication protein (272 aa; *E. coli* UTI89 [uropathogenic strains])	YP 544031	7.20E-151 (271/294) 92.2%	DS

47	38417→39115	232	52.79	Dam	DNA adenine methylase (217aa; *E. coli* UTI89 [uropathogenic strains])	YP 544033	1.8E-126 (216/217) 99.5%	DS
48	39112→39438	108	55.96	LexA_DNA_bind (RecA-mediated autopeptidases)	SOS-response transcriptional repressors (108aa; *E. coli* 101-1 [EAEC])	ZP 00925782	1.20E-54 (108/108) 100.0%	DS
49	39435→39824	129	55.64	Endodeoxy-ribonuclease (RusA)	Holliday junction resolvase (129aa; *E. coli* B171 [EPEC O111:NM])	ZP 00712000	9.60E-70 (129/129) 100%	DS
50	39844→40653	269	50.99	Unknown	Hypothetical protein EcolB_01000946 (KilA-N domain) (269 aa; *E. coli* B171 [EPEC O111:NM])	ZP 00712001	8.90E-150 (267/269) 99.3%	DS
51	40733→41650	305	55.77	Unknown	Hypothetical protein YdfU (331 aa; *E. coli* UTI89 [uropathogenic strains])	YP 544037	4.60E-180 (302/305) 99.0%	1
52	41668→42030	120	57.85	Q protein	Putative phage antiterminator (122 aa; *Salmonella typhimurium* LT2)	AAL21141	2.40E-52 (98/120) 81.7%	13
53	42054←42497	147	32.21	Unknown	Unknown			
54	42501←43442	313	35.35	Unknown	Unknown			
55	43697→43924	75	49.56	Unknown	Prophage CP-933O conserved hypothetical protein (75 aa; *E. coli* UTI89 [uropathogenic strains])	YP 544040	7.90E-35 (72/75) 96.0%	1
56	44075→45127	350	47.86	DNA methylase	DNA modification methylase (350 aa; *E. coli* B171 [EPEC O111:NM])	ZP 00709127	0 (349/350) 99.7%	DS
57	45194→45409	71	46.76	Lysis protein	Predicted phage lysis protein (71 aa; *E. coli* K12 DLP12 prophage)	AAC73655	8.50E-34 (71/71) 100.0%	14
58	45409→45906	165	50.8	Lysin	Lysozyme-like protein (165aa; *Shigella sonnei* Ss046)	YP 311635	2.40E-91 (162/165) 98.2%	4
59	45903→46370	155	54.49	Lysin (Rz)	Chromosome segregation ATPases (155 aa; *E. coli* B171 [EPEC O111:NM])	ZP 00712006	3.60E-80 (154/155) 99.4%	DS
60	46392→46754	120	49.86	Unknown	Hypothetical protein EcolB_01000953 (KilA-N) (127 aa; *E. coli* B171 [EPEC O111:NM])	ZP 00712007	3.20E-65 (118/120) 98.3%	DS

[a] DS: Direct submission

References

1. Chen SL, Hung C, Xu J, Reigstad CS, Magrini V, Sabo A, Blasiar D, Bieri T, Meyer RR, Ozersky P, *et al*. (2006) *Proc Natl Acad Sci USA* 103:5977-5982.
2. Yokoyama K, Makino K, Kubota Y, Watanabe M, Kimura S, Yutsudo CH, Kurokawa K, Ishii K, Hattori M, Tatsuno I, *et al*. (2000) *Gene* 258:127-39.
3. Hayashi T, Makino K, Ohnishi M, Kurokawa K, Ishii K, Yokoyama K, Han CG, Ohtsubo E, Nakayama K, Murata T, *et al*. (2001) *DNA Res* 8:11-22.
4. Yang F, Yang J, Zhang X, Chen L, Jiang Y, Yan Y, Tang X, Wang J, Xiong Z, Dong J, *et al.* (2005) *Nucleic Acids Res* 33:6445-6458.
5. Perna N. T, Plunkett G 3rd, Burland V, Mau B, Glasner JD, Rose DJ, Mayhew GF, Evans PS, Gregor J, Kirkpatrick HA, *et al*. (2001) *Nature* 409:529-533.
6. Sato T, Shimizu T, Watarai M, Kobayashi M, Kano S, Hamabata T, Takeda Y, Yamasaki S (2003) *Gene* 309:35-48.
7. Tobe T, Beatson SA, Taniguchi H, Abe H, Bauley CM, Fivian A, Younis R, Matthews S, Marches O, Frankel G, *et al.* (2006) *Proc Natl Acad Sci USA* 103:14941-14946.
8. Marches O, Ledger TN, Boury M, Ohara M, Tu X, Goffaux F, Mainil J, Rosenshine I, Sugai M, De Rycke J, *et al*. (2003) *Mol Microbiol* 50:1553-1567.
9. Pickett CL, Cottle DL, Pesci EC, Bikah G (1994) *Infect Immun* 62:1046-1051.
10. Huan PT, Whittle BL, Bastin DA, Lindberg AA, Verma NK (1997) *Gene* 195:207-216.
11. Allison GE, Angeles D, Tran-Dinh N, Verma NK (2002) *J Bacteriol* 184:1974-1987.
12. Welch RA, Burland V, Plunkett G 3rd, Redford P, Roesch P, Rasko DA, Buckles EL, Liou S-R, Boutin A, Hackett J, *et al*. (2002) *Proc Natl Acad Sci USA* 99:17020-17024.
13. McClelland M, Sanderson KE, Spieth J, Clifton SW, Latreille P, Courtney L, Porwollik S, Ali J, Dante M, Du F, *et al.* (2001) *Nature* 413:852-856.
14. Blattner FR, Plunkett G, Bloch CA, Perna NT, Burland V, Riley M, Collado-Vides J, Glasner JD, Rode CK, Mayhew GF, *et al.* (1997) *Science* 277:1453-1474.