

**Supporting information to: Evidence of *in vivo* prophage induction during *Clostridium difficile* infection**

Mathieu Meessen-Pinard, Ognjen Sekulovic, Louis-Charles Fortier\*

Département de microbiologie et d'infectiologie, Faculté de médecine et des sciences de la santé,  
Université de Sherbrooke, Québec, Canada, J1E 4K8

\*Corresponding author:

Louis-Charles Fortier. Département de microbiologie et d'infectiologie, Faculté de médecine et des sciences de la santé, Université de Sherbrooke, 3201 rue Jean-Mignault, Sherbrooke, Québec, Canada, J1E 4K8. Phone: (819) 821-8000 ext 75322. Fax: (819) 820-6831. email: [Louis-Charles.Fortier@USherbrooke.ca](mailto:Louis-Charles.Fortier@USherbrooke.ca)

Keywords: *Clostridium difficile*, temperate bacteriophage, prophage, spontaneous induction

Running title: *In vivo* prophage induction in *C. difficile*

M.M.-P. and O.S. contributed equally to this work.

**TABLE S1. Putative coding sequences from  $\phi$ MMP02 and bioinformatics analyses of the deduced proteins**

ORF	Start	End	Size (pb)	Size (aa)	Ribosome binding site (RBS)	Predicted fonction <sup>a</sup>	BLASTp hit - accession number [source organism] <sup>b</sup>	Alignement (%ID)	Size (aa) <sup>c</sup>	E value
1	62	763	760	233	aaacgagg <sub>ggggaat</sub> ATG	Terminase small subunit	YP_002290877 putative terminase small subunit [ <i>Clostridium</i> phage phiCD27]	232/233 (99%)	233	2E-131
2	756	2165	1410	469	aaacgtgg <sub>ggttaacc</sub> ATG	Terminase large subunit	YP_002290878 putative terminase B [ <i>Clostridium</i> phage phiCD27]	464/469 (99%)	469	0E+00
3	2156	3658	1503	500	gaaaggggg <sub>gaggtgt</sub> ATG	Portal protein	YP_002290879 putative portal protein [ <i>Clostridium</i> phage phiCD27]	491/500 (99%)	500	0
4	3645	4658	1014	337	aaaggga <sub>aaan</sub> ATG	Head morphogenesis protein	YP_002290880 putative head morphogenesis protein [ <i>Clostridium</i> phage phiCD27]	331/335 (99%)	335	0
5	4678	6012	1335	444	gaaaggtgg <sub>tgattaa</sub> ATG		YP_002290881 hypothetical protein phiCD27_gp05 [ <i>Clostridium</i> phage phiCD27]	443/444 (99%)	444	0
6	6028	6411	384	127	aaaggag <sub>tgatgttat</sub> ATG		YP_002290882 hypothetical protein phiCD27_gp06 [ <i>Clostridium</i> phage phiCD27]	127/127 (100%)	127	2E-67
7	6430	7467	1038	345	aaaggag <sub>tgataaata</sub> ATG		YP_002290883 hypothetical protein phiCD27_gp07 [ <i>Clostridium</i> phage phiCD27]	345/345 (100%)	345	0
8	7507	7845	339	112	gagggtg <sub>aaat</sub> ATG		YP_002290884 hypothetical protein phiCD27_gp08 [ <i>Clostridium</i> phage phiCD27]	111/112 (99%)	112	1E-57
9	7842	8219	378	125	agaa <sub>ttgaaat</sub> ATG		YP_002290885 hypothetical protein phiCD27_gp09 [ <i>Clostridium</i> phage phiCD27]	123/125 (99%)	125	4E-64
10	8221	8631	411	136	gaaagga <sub>tgattaa</sub> ATG	Putative tail component protein	YP_002290886 hypothetical protein phiCD27_gp10 [ <i>Clostridium</i> phage phiCD27]	137/139 (99%)	148	4E-75
11	8632	9453	822	273	ggag <sub>tgattaa</sub> ATG		YP_002290887 hypothetical protein phiCD27_gp11 [ <i>Clostridium</i> phage phiCD27]	262/273 (96%)	273	4E-151
12	9466	9669	204	67	aggagg <sub>ggcagaa</sub> TTG		YP_002290888 hypothetical protein phiCD27_gp12 [ <i>Clostridium</i> phage phiCD27]	60/67 (90%)	67	2E-24
13	9671	11092	1422	473	agaaagagagg <sub>gaaataa</sub> ATG	Tail sheath protein (XkdK-related)	YP_002290889 hypothetical protein phiCD27_gp13 [ <i>Clostridium</i> phage phiCD27]	454/473 (96%)	473	0
14	11108	11521	414	137	aaaggagg <sub>taagaa</sub> ATG	XkdM-related protein	YP_002290890 hypothetical protein phiCD27_gp14 [ <i>Clostridium</i> phage phiCD27]	57/136 (42%)	140	1E-24
15	11605	12096	492	163	ggagg <sub>aaatgt</sub> ATG	XkdN-related protein	YP_002290891 hypothetical protein phiCD27_gp15 [ <i>Clostridium</i> phage phiCD27]	87/145 (60%)	157	7E-43
16	12102	12254	153	50						
17	12273	15547	4275	1424	agaaagg <sub>aaag</sub> CTG	Phage tail tape measure protein	YP_002290892 putative phage tail tape measure protein [ <i>Clostridium</i> phage phiCD27]	641/1527 (42%)	1429	0
18	16630	17373	744	247	aggggg <sub>aaatrr</sub> ATG	Putative secreted protein	YP_002290893 hypothetical protein phiCD27_gp17 [ <i>Clostridium</i> phage phiCD27]	31/58 (54%)	285	7E-07
19	17442	18113	672	223	aaaagagg <sub>tgataac</sub> TTG	Peptidoglycan-binding LysM protein	YP_002290894 putative peptidoglycan-binding LysM [ <i>Clostridium</i> phage phiCD27]	209/223 (94%)	223	4E-111
20	18110	19072	963	320	aaagtgg <sub>tgataaccttaac</sub> ATG	XkdQ-related protein	YP_002290895 hypothetical protein phiCD27_gp19 [ <i>Clostridium</i> phage phiCD27]	309/320 (97%)	320	9E-171
21	19065	19430	366	121	aaggagg <sub>gaaattaccttaata</sub> ATG		YP_002290896 hypothetical protein phiCD27_gp20 [ <i>Clostridium</i> phage phiCD27]	110/120 (92%)	120	1E-57
22	19405	19863	459	152	ggagg <sub>gattgaa</sub> ATG	XkdS-related protein	YP_002290897 hypothetical protein phiCD27_gp21 [ <i>Clostridium</i> phage phiCD27]	142/152 (94%)	152	1E-76
23	19872	21008	1137	378	aaagtagg <sub>tgattat</sub> ATG	Baseplate J family protein	YP_002290898 hypothetical protein phiCD27_gp22 [ <i>Clostridium</i> phage phiCD27]	343/378 (91%)	377	0
24	21005	21634	630	209	aggtgg <sub>tggtg</sub> ATG	XkdT-related protein	YP_001254823 hypothetical protein CBO2332 [ <i>C. botulinum</i> A str. ATCC 3502]	116/209 (56%)	211	8E-64
25	21647	22669	1023	340	aggagg <sub>gggtataa</sub> ATG	Phage tail fiber protein	YP_001110744 putative tail fiber protein [ <i>Clostridium</i> phage phiC2]	253/338 (75%)	341	1E-138
26	22684	24366	1683	560	aaaggagag <sub>tgagaaa</sub> ATG		ZP_05352971 hypothetical protein CdifA_19608 [ <i>C. difficile</i> ATCC 43255]	456/570 (80%)	571	0
27	24383	24676	294	97	gagggtg <sub>aaana</sub> ATG		YP_001087449 hypothetical protein CD0968 [ <i>C. difficile</i> 630]	86/97 (89%)	97	4E-42
28	24676	24858	183	60	aggaggaa <sub>attata</sub> ATG		ZP_05352945 hypothetical protein CdifA_19478 [ <i>C. difficile</i> ATCC 43255]	57/59 (97%)	60	1E-23
29	24891	25829	939	312	gaaagaag <sub>tgactaa</sub> ATG		YP_002290905 hypothetical protein phiCD27_gp29 [ <i>Clostridium</i> phage phiCD27]	304/310 (99%)	410	5E-175
30	26119	26418	300	99	gaaaggag <sub>aatcat</sub> ATG		YP_002290906 hypothetical protein phiCD27_gp30 [ <i>Clostridium</i> phage phiCD27]	94/99 (95%)	99	9E-47
31	26420	26557	138	45	aatggg <sub>ggtatagt</sub> ATG	XkdX-related protein	YP_002290907 hypothetical protein phiCD27_gp31 [ <i>Clostridium</i> phage phiCD27]	43/45 (96%)	45	3E-16
32	26589	26819	231	76	aggagg <sub>tttc</sub> ATG		YP_002290908 hypothetical protein phiCD27_gp32 [ <i>Clostridium</i> phage phiCD27]	76/76 (100%)	76	7E-36
33	26838	27095	258	85	aggagg <sub>aaataaaga</sub> ATG	Holin	YP_002290909 holin [ <i>Clostridium</i> phage phiCD27]	85/85 (100%)	85	7E-40
34	27095	27907	813	270	aggagg <sub>aaataaata</sub> ATG	Endolysin	EHJ29023.1 N-acetylmuramoyl-L-alanine amidase [ <i>Clostridium difficile</i> 050-P50-2011]	246/270 (92%)	270	2E-142
35	28157	28648	492	163	aggggg <sub>attaaa</sub> ATG		ZP_01994810 hypothetical protein DORLON_00799 [ <i>Dorea longicatena</i> DSM 13814]	43/184 (24%)	185	5E-08
36	28650	29216	567	188	aggaaag <sub>aatagt</sub> ATG	Putative membrane protein	YP_004062980 hypothetical protein CKC_03715 [ <i>C. Liberibacter</i> CLso-ZC1]	32/91 (36%)	158	2E-05
37	29815	29627	189	62	gaaagga <sub>ataaaaat</sub> ATG		YP_001110759 hypothetical protein phiC2p43 [ <i>Clostridium</i> phage phiC2]	62/62 (100%)	62	3E-26
38	30075	30218	144	47	aggaaggagg <sub>ggagaa</sub> ATG		YP_001089402.1 hypothetical protein phiCD27_gp41 [ <i>Clostridium</i> phage phiCD27]	47/47 (100%)	47	2E-24
39	30726	31115	390	129	aaggag <sub>tgacataa</sub> ATG	Regulatory protein	ZP_05349389.1 putative phage regulatory protein [ <i>Clostridium difficile</i> ATCC 43255]	65/92 (71%)	93	3E-31
40	31208	31489	282	93	aaaggggg <sub>tgacataa</sub> ATG	Regulatory protein	ZP_05349390 putative regulatory protein [ <i>C. difficile</i> ATCC 43255]	83/89 (94%)	126	8E-42
41	32850	31804	1047	348	ggggg <sub>gtctaaa</sub> ATG	Integrase	ZP_05349367 phage integrase family protein [ <i>C. difficile</i> ATCC 43255]	306/348 (88%)	351	0
42	33142	32927	216	71	aaaggtgg <sub>attataaa</sub> ATG	Lambda repressor-like	ZP_05349368 hypothetical protein CdifA_01317 [ <i>C. difficile</i> ATCC 43255]	56/60 (94%)	71	5E-23

**TABLE S1. Putative coding sequences from φMMP02 and bioinformatics analyses of the deduced proteins**

ORF	Start	End	Size (pb)	Size (aa)	Ribosome binding site (RBS)	Predicted fonction <sup>a</sup>	BLASTp hit - accession number [source organism] <sup>b</sup>	Alignement (%ID)	Size (aa) <sup>c</sup>	E value
43	33428	33559	132	43	agggga <sub>13</sub> ggttat <sup>-</sup> TTG					
44	33997	33576	402	133	aggtgagg <sub>13</sub> gfgaan <sup>-</sup> ATG	Lambda repressor-like	ZP_04088144.1 Transcriptional regulator, MerR [Bacillus thuringiensis BGSC 4BD1]	42/77 (55%)	153	2E-16
45	34409	34504	96	31	aaaaggag <sub>13</sub> caagtat <sup>-</sup> ATG					
46	34867	34649	219	72	ggag <sub>13</sub> cggtgatt <sup>-</sup> ATG	Phage repressor (XRE family)	YP_529597 hypothetical protein CDBPCV119_gp46 [Clostridium phage phi CD119]	70/72 (98%)	72	8E-44
47	34980	35174	195	64	aaaggag <sub>13</sub> agg <sup>-</sup> GTG	Phage repressor	YP_529597 hypothetical protein CDBPCV119_gp46 [Clostridium phage phi CD119]	26/61 (43%)	72	2E-08
48	35216	36010	795	264	agaaaaggag <sub>13</sub> ctttgat <sup>-</sup> ATG	Phage antirepressor	YP_529598 putative antirepressor [Clostridium phage phi CD119]	245/262 (94%)	277	2E-140
49	36094	36474	381	126	agggag <sub>13</sub> daaatac <sup>-</sup> ATG		YP_001110771 hypothetical protein phiC2p56 [Clostridium phage phiC2]	122/126 (97%)	126	1E-61
50	36613	36464	150	49	aaag <sub>13</sub> g <sup>-</sup> GTG		YP_001110772 Hypothetical protein phiC2p57 [Clostridium phage phiC2]	49/50 (98%)	50	2E-17
51	36675	36872	198	65	aaaggagg <sub>13</sub> gaaatg <sup>-</sup> ATG	Phage regulator	YP_001110773 putative repressor [Clostridium phage phiC2]	62/65 (96%)	65	4E-27
52	36911	37117	207	68	aaagttag <sub>13</sub> ggttatg <sup>-</sup> ATG		ZP_05349377 hypothetical protein CdifA_01362 [C. difficile ATCC 43255]	32/63 (51%)	74	9E-10
53	37188	37292	105	34	aggaaag <sub>13</sub> gggattg <sup>-</sup> TTG		ZP_05349378 hypothetical protein CdifA_01367 [C. difficile ATCC 43255]	30/34 (89%)	34	5E-09
54	37530	37847	318	105	aagagggg <sub>13</sub> attaac <sup>-</sup> ATG		ZP_05349380 hypothetical protein CdifA_01377 [C. difficile ATCC 43255]	97/105 (93%)	105	2E-44
55	37930	38439	510	169	aaatggagg <sub>13</sub> gaaat <sup>-</sup> ATG		YP_001110779 Gp157-like protein [Clostridium phage phiC2]	163/169 (97%)	169	2E-84
56	38449	39057	609	202	aaaagtag <sub>13</sub> ggagacc <sup>-</sup> ATG	Putative ERF protein	YP_529602 hypothetical protein CDBPCV119_gp51 [Clostridium phage phi CD119]	199/202 (99%)	202	1E-110
57	39067	39948	882	293	aagag <sub>13</sub> taggtgagtg <sup>-</sup> TTG	Phage replication protein	YP_001110781 putative DnaD [Clostridium phage phiC2]	280/293 (96%)	296	4E-156
58	40009	40182	174	57	agaaaggggg <sub>13</sub> gataaa <sup>-</sup> ATG		ZP_05349384 putative phage resolvase/integrase (partial) [C. difficile ATCC 43255]	57/57 (100%)	57	4E-24
59	40198	40614	417	138	agaaggggg <sub>13</sub> ccatgttaa <sup>-</sup> ATG	SSB	YP_529604 hypothetical protein CDBPCV119_gp53 [Clostridium phage phi CD119]	115/138 (84%)	135	3E-60
60	40685	41098	414	137	ggagg <sub>13</sub> gattgaa <sup>-</sup> ATG	YopX family protein	ZP_05349386 hypothetical protein CdifA_01407 [C. difficile ATCC 43255]	133/137 (98%)	136	5E-69
61	41085	41336	252	83	gagggggg <sub>13</sub> catagaa <sup>-</sup> ATG		ZP_05349387 hypothetical protein CdifA_01412 [C. difficile ATCC 43255]	76/83 (92%)	83	3E-36
62	41354	41707	354	117	agggagggg <sub>13</sub> ggaactaaat <sup>-</sup> ATG		YP_529608.1 hypothetical protein CDBPCV119_gp57 [Clostridium phage phi CD119]	48/109 (44%)	106	1E-15
63	41676	41816	141	46	aaagag <sub>13</sub> gagcag <sup>-</sup> ATG		ZP_05353027 hypothetical protein CdifA_19888 [C. difficile ATCC 43255]	42/44 (96%)	46	8E-14
64	41813	41929	117	38	gagggga <sub>13</sub> gittagaana <sup>-</sup> ATG		ZP_05332132 hypothetical protein CdifQCD-6_20253 [C. difficile QCD-63q42]	38/38 (100%)	38	2E-13
65	41932	42108	177	58	aaagagg <sub>13</sub> ggtataaga <sup>-</sup> ATG		ZP_05332133 hypothetical protein CdifQCD-6_20258 [C. difficile QCD-63q42]	58/58 (100%)	58	1E-23
66	42218	43291	1074	357	aaggagg <sub>13</sub> cggtgta <sup>-</sup> TTG	DNA-cytosine methyltransferase	YP_001110788 putative cytosine-C5 specific DNA methylase [C. difficile phage phiC2]	311/311 (100%)	316	0
67	43371	43580	210	69	aagggg <sub>13</sub> gggataa <sup>-</sup> ATG		YP_001110789 hypothetical protein phiC2p74 [Clostridium phage phiC2]	68/69 (99%)	69	2E-30
68	43592	44596	1005	334	aaggggtgaaatt <sup>-</sup> ATG		YP_001110790 hypothetical protein phiC2p75 [Clostridium phage phiC2]	315/334 (95%)	334	3E-174
69	44605	44772	168	55	gaagaaaag <sub>13</sub> aggtgacat <sup>-</sup> ATG		YP_529616 hypothetical protein CDBPCV119_gp65 [Clostridium phage phi CD119]	44/52 (85%)	57	6E-17
70	44857	45129	273	90	aggaaggagg <sub>13</sub> gagatt <sup>-</sup> ATG		ZP_05332140 hypothetical protein CdifQCD-6_20293 [C. difficile QCD-63q42]	72/89 (81%)	95	7E-32
71	45152	45382	231	76	aaaggg <sub>13</sub> ggtgactt <sup>-</sup> ATG	ASCH domain protein	ZP_05332243.1 hypothetical protein CdifQCD-6_20810 [C. difficile QCD-63q42]	41/75 (55%)	140	2E-21
72	45434	45934	501	166	gaagaaga <sub>13</sub> gcagttggt <sup>-</sup> GTG		YP_001089447 hypothetical protein CD2932 [C. difficile 630]	90/132 (69%)	166	3E-40
73	45924	46367	444	147	aaatgagg <sub>13</sub> ingccat <sup>-</sup> ATG	RusA endonuclease	YP_001089446 phage protein [C. difficile 630]	82/147 (56%)	139	5E-40
74	46324	46566	243	80	aagaaa <sub>13</sub> attata <sup>-</sup> GTG					
75	46671	47177	507	168	aggag <sub>13</sub> ggtgaggt <sup>-</sup> ATG	Putative sigma factor	ZP_03292416 hypothetical protein CLOHIR_00359 [C. hiranonis DSM 13275]	53/154 (35%)	172	1E-09
76	47718	48395	678	225	aggagg <sub>13</sub> tagtgr <sup>-</sup> ATG	Phage repressor (Bro-N superfamily)	YP_002290951 hypothetical protein phiCD27_gp75 [Clostridium phage phiCD27]	225/225 (100%)	225	1E-127

<sup>a</sup> Inferred from BLASTp analyses on GenBank and ACLAME databases, and by conserved domains found by InterProScan analysis and through the conserved domains database (CDD) at NCBI

<sup>b</sup> Only the best or the most relevant BLASTp hits are shown

<sup>c</sup> Size in amino acids (aa) of the BLASTp database subject

ERF = Essential Recombination Function protein; SSB = Single-Stranded DNA-Binding protein

TABLE S2. Putative coding sequences from φMMP04 and bioinformatics analyses of the deduced proteins

ORF	Start	End	Size (pb)	Size (aa)	Ribosome binding site (RBS)	Predicted fonction <sup>a</sup>	BLAST hit and accession number [source organism] <sup>b</sup>	Alignment (%ID)	Size (aa) <sup>c</sup>	E value
1	22	1671	1650	189	aaaggtaga <sub>cttaagaa</sub> ATG	Terminase	EHJ28459 putative phage terminase, large subunit [ <i>Clostridium difficile</i> 002-P50-2011]	504/549 (92%)	549	0
2	1687	2841	1155	384	aaggggg <sub>tgaaaa</sub> ATG	Phage portal protein	EHJ28460 phage portal protein [ <i>Clostridium difficile</i> 002-P50-2011]	359/384 (93%)	384	0
3	2847	3584	738	245	ggaaggagg <sub>cgagaaattaa</sub> ATG	ATP-dependent Clp protease	EHJ28461 endopeptidase Clp [ <i>Clostridium difficile</i> 002-P50-2011]	214/245 (87%)	245	7E-149
4	3577	4845	1269	422	aggggg <sub>cttaa</sub> ATG	Major capsid protein	EHJ28462 phage capsid family protein [ <i>Clostridium difficile</i> 002-P50-2011]	184/433 (42%)	422	7E-87
5	4855	5142	288	95	gaagtagg <sub>gttttga</sub> ATG	Head-Tail connector (gp6-like)	EHJ28463 phage DNA packaging protein [ <i>Clostridium difficile</i> 002-P50-2011]	91/95 (96%)	95	2E-47
6	5144	5482	339	112	agtagg <sub>actaaa</sub> ATG	Phage head-tail adaptor	EHJ28464 putative phage head-tail adaptor [ <i>Clostridium difficile</i> 002-P50-2011]	106/112 (95%)	112	8E-71
7	5475	5888	414	137	aaagaagg <sub>tgttaca</sub> ATG		EHJ28465 phage protein, HK97 gp10 family [ <i>Clostridium difficile</i> 002-P50-2011]	128/137 (93%)	137	1E-33
8	5900	6316	417	138	agaggggg <sub>caaga</sub> GTG		EHJ28466 hypothetical protein HMPREF1122_02772 [ <i>Clostridium difficile</i> 002-P50-2011]	129/138 (93%)	138	1E-84
9	6334	7401	1068	355	aaaggagg <sub>ctttatt</sub> TTG	Phage tail sheath protein (XkdK-related)	EHJ28467 hypothetical protein HMPREF1122_02773 [ <i>Clostridium difficile</i> 002-P50-2011]	339/355 (95%)	355	0
10	7414	7692	279	92	aagaagg <sub>tgatag</sub> ATG	XkdM-related protein	EHJ28468 hypothetical protein HMPREF1122_02774 [ <i>Clostridium difficile</i> 002-P50-2011]	75/87 (86%)	149	2E-42
11	7928	8293	366	121	ggaaaa <sub>taaaagaa</sub> ATG	XkdN-related protein	EHJ28469 phage XkdN-like protein [ <i>Clostridium difficile</i> 002-P50-2011]	121/121 (100%)	121	1E-80
12	8311	8484	174	57	aaaa <sub>cgatgtgaa</sub> ATG		EHJ28470 hypothetical protein HMPREF1122_02776 [ <i>Clostridium difficile</i> 002-P50-2011]	57/57 (100%)	57	3E-37
13	8477	10246	1770	589	ggagg <sub>tgtaaaa</sub> ATG	Tail tape measure protein	EHJ28471 hypothetical protein HMPREF1122_02777 [ <i>Clostridium difficile</i> 002-P50-2011]	570/579 (98%)	584	0
14	10230	10865	636	211	gaaaatgggg <sub>tggttca</sub> ATG	LysM (XkdP-related)	EHJ28472 LysM domain protein [ <i>Clostridium difficile</i> 002-P50-2011]	206/211 (98%)	211	4E-151
15	11105	11383	279	92	aaagaaa <sub>ctctatgagaa</sub> ATG	Cell wall hydrolase (XkdQ-related)	YP_529574 XkdQ protein [ <i>Clostridium</i> phage phi CD119]	83/90 (92%)	228	8E-41
16	11380	12624	1245	414	aaaaaaag <sub>at</sub> ATG	Cell wall hydrolase (NlpC/P60 family)	EHJ28473 NlpC/P60 family protein [ <i>Clostridium difficile</i> 002-P50-2011]	402/414 (97%)	505	0
17	12639	12965	327	35	aaaggtgg <sub>tgtaaa</sub> GTG		YP_001087867 hypothetical protein CD1369 [ <i>C. difficile</i> 630]	101/108 (94%)	108	0
18	12965	13393	429	142	gggtgag <sub>cttata</sub> ATG	XkdS-related protein	EHJ28838 phage protein XkdS [ <i>Clostridium difficile</i> 002-P50-2011]	133/142 (94%)	142	1E-91
19	13386	14438	1053	350	ggggagg <sub>tgaaagta</sub> ATG	Baseplate j-like (XkdT-related)	YP_529578 XkdT protein [ <i>Clostridium</i> phage phi CD119]	315/350 (90%)	350	0
20	14438	15055	618	205	gaggtgg <sub>tgattta</sub> ATG		YP_529579 hypothetical protein CDBPCV119_gp28 [ <i>Clostridium</i> phage phi CD119]	172/202 (85%)	202	9E-99
21	15056	15838	783	73	aggagg <sub>gatata</sub> TTG	Tail fiber protein	ZP_05332079 putative phage tail fiber protein [ <i>C. difficile</i> QCD-63q42]	226/260 (87%)	260	2E-120
22	15855	17120	1266	421	agaaagtgagg <sub>tgtaaa</sub> GTG		ZP_05354705 hypothetical protein CdifQCD-7_02151 [ <i>C. difficile</i> QCD-76w55]	249/413 (60%)	495	3E-122
23	17340	17633	294	97	aaaagagg <sub>tgataaa</sub> ATG		YP_001087449 hypothetical protein CD0968 [ <i>C. difficile</i> 630]	91/97 (94%)	97	2E-45
24	17633	17815	183	60	aggagg <sub>ctatata</sub> ATG		YP_529583 hypothetical protein CDBPCV119_gp32 [ <i>Clostridium</i> phage phi CD119]	57/60 (95%)	60	7E-25
25	17877	18308	432	143	aaggagg <sub>aaat</sub> ATG	Holin	ZP_05331982 putative cell wall hydrolase protein [ <i>C. difficile</i> QCD-63q42]	138/143 (97%)	143	1E-73
26	18308	19156	849	283	aaaggagg <sub>tgtaataa</sub> ATG	Endolysine	YP_002290910 endolysin [ <i>Clostridium</i> phage phiCD27]	232/283 (82%)	270	1E-126
27	19640	19281	360	119	aaaggggaa <sub>tgataat</sub> ATG		YP_004306132 hypothetical protein phiCD6356_31 [ <i>Clostridium</i> phage phiCD6356]	109/119 (92%)	129	5E-53
28	20460	19642	819	272	aaaaagggg <sub>tgatttaa</sub> TTG	ParA-like protein	YP_004306133 ParA-like protein [ <i>Clostridium</i> phage phiCD6356]	249/272 (92%)	272	2E-143
29	20993	20811	183	60	aaaggag <sub>taaaga</sub> ATG		YP_004306134 hypothetical protein phiCD6356_33 [ <i>Clostridium</i> phage phiCD6356]	57/60 (95%)	60	2E-25
30	21361	21233	129	42	agaaaggagg <sub>caagaga</sub> ATG	Membrane / Secreted protein	ZP_05331937 hypothetical protein CdifQCD-6_19278 [ <i>C. difficile</i> QCD-63q42]	41/42 (98%)	42	4E-07
31	21946	21794	153	50	gaagggagg <sub>gaatagt</sub> ATG	Membrane / Secreted protein	YP_004306135 hypothetical protein phiCD6356_34 [ <i>Clostridium</i> phage phiCD6356]	49/50 (98%)	50	1E-18
32	22171	22067	105	34	gaggaa <sub>taaga</sub> TTG					
33	22787	22254	534	177	aaaggggaa <sub>tgactt</sub> GTG	DNA polymerase III, alpha subunit	ZP_03292395 hypothetical protein CLOHIR_00338 [ <i>C. hiranonis</i> DSM 13275]	79/129 (61%)	276	4E-34
34	23222	25378	2157	542	ggagg <sub>ggataa</sub> ATG	DNase / helicase	ZP_05331934 putative deoxyribonuclease [ <i>C. difficile</i> QCD-63q42]	639/721 (89%)	721	0
35	25399	25737	339	112	ggagg <sub>aatgaat</sub> ATG		ZP_05331933 hypothetical protein CdifQCD-6_19258 [ <i>C. difficile</i> QCD-63q42]	62/114 (54%)	115	7E-22
36	26078	26185	108	35	agggggg <sub>ggtttg</sub> TTG		YP_004508416 hypothetical protein phiCD38-2_gp38 [ <i>Clostridium</i> phage phiCD38-2]	19/30 (63%)	31	2E-04
37	26589	26215	375	124	gaaaggg <sub>ttgtat</sub> ATG	Transcriptional regulator (XRE family)	ZP_05331920 transcriptional regulator, xre family protein [ <i>C. difficile</i> QCD-63q42]	77/117 (66%)	117	2E-37
38	27011	27436	426	141	gaagggaaaagg <sub>tgatatt</sub> ATG		YP_004508419 hypothetical protein phiCD38-2_gp41 [ <i>Clostridium</i> phage phiCD38-2]	43/138 (31%)	141	8E-11
39	27547	27696	150	49	aaggag <sub>maacaa</sub> ATG		[YP_004306146 hypothetical protein phiCD6356_45 [ <i>Clostridium</i> phage phiCD6356]	25/42 (60%)	58	3E-05
40	27710	28063	354	117	gaagggg <sub>tgtttga</sub> TTG		YP_004306147 hypothetical protein phiCD6356_46 [ <i>Clostridium</i> phage phiCD6356]	35/99 (35%)	121	5E-04
41	28075	28248	174	57	aaggggg <sub>ctattt</sub> ATG					

TABLE S2. Putative coding sequences from φMMP04 and bioinformatics analyses of the deduced proteins

ORF	Start	End	Size (pb)	Size (aa)	Ribosome binding site (RBS)	Predicted fonction <sup>a</sup>	BLAST hit and accession number [source organism] <sup>b</sup>	Alignement (%ID)	Size (aa) <sup>c</sup>	E value
42	28245	28349	105	34	-					
43	28342	28599	258	85	agaagggaag <sub>tgtagaa</sub> ATG		YP_004306149 hypothetical protein phiCD6356_48 [ <i>Clostridium</i> phage phiCD6356]	53/71 (75%)	71	7E-23
44	28575	28727	153	50	agaaaagga <sub>ggggcaaa</sub> ATG		YP_004306152 hypothetical protein phiCD6356_51 [ <i>Clostridium</i> phage phiCD6356]	25/42 (60%)	49	0.67
45	28717	28977	261	86	agaaagagagg <sub>ttcaaaag</sub> ATG		YP_004306155 hypothetical protein phiCD6356_54 [ <i>Clostridium</i> phage phiCD6356]	52/73 (71%)	86	1E-18
46	29221	29688	468	155	aaaggggg <sub>taaaattg</sub> TTG		YP_004306157 hypothetical protein phiCD6356_56 [ <i>Clostridium</i> phage phiCD6356]	67/156 (43%)	153	3E-17
47	29933	30436	504	167	agaaa <sub>ctatggtatttt</sub> ATG	Integrase	EHJ28455 site-specific recombinase, phage integrase family [ <i>Clostridium difficile</i> 002-P50-2011]	124/167 (74%)	200	8E-67
48	30634	30924	291	96	aaaaagagg <sub>ggitta</sub> ATG		YP_659579 hypothetical protein CDP03 [ <i>C. difficile</i> 630]	45/67 (67%)	71	2E-17
49	30924	31217	294	97	agagaa <sub>gggtatatt</sub> ATG	HNH endonuclease	YP_659578 putative HNH endonuclease [ <i>C. difficile</i> 630]	93/97 (96%)	97	4E-41
50	31312	31425	114	37	gaaaaa <sub>tcaga</sub> ATG		EHJ28458 hypothetical protein HMPREF1122_02764 [ <i>Clostridium difficile</i> 002-P50-2011]	35/35 (100%)	129	1E-13
51	31430	25	270	90	agaaaagaaa <sub>gtaaaatt</sub> TTG		EHJ28458 hypothetical protein HMPREF1122_02764 [ <i>Clostridium difficile</i> 002-P50-2011]	83/89 (93%)	129	1E-49

<sup>a</sup> Inferred from BLASTp analyses on GenBank and ACLAME databases, and by conserved domains found by InterProScan analysis and through the conserved domains database (CDD) at NCBI

<sup>b</sup> Only the best or most relevant BLASTp hits are shown

<sup>c</sup> Size in amino acids (aa) of the BLASTp database subject