

**Complete genome sequence of the newly discovered temperate *Clostridioides difficile* bacteriophage phiCDKH01 of the family *Siphoviridae***

Archives of Virology

Krzysztof Hinc<sup>1\*</sup>, Monika Kabała<sup>2</sup>, Adam Iwanicki<sup>1</sup>, Gajane Martirosian<sup>2</sup>, Alessandro Negri<sup>1</sup> Michał Obuchowski<sup>1</sup>

<sup>1</sup>Department of Molecular Bacteriology, Intercollegiate Faculty of Biotechnology of University of Gdańsk and Medical University of Gdańsk, Dębinki 1, Gdańsk 80-211, Poland

<sup>2</sup>Department of Medical Microbiology, Medical University of Silesia, Faculty of Medical Sciences in Katowice, Medyków 18, Katowice 40-752, Poland

\* corresponding author: krzysztof.hinc@biotech.ug.edu.pl

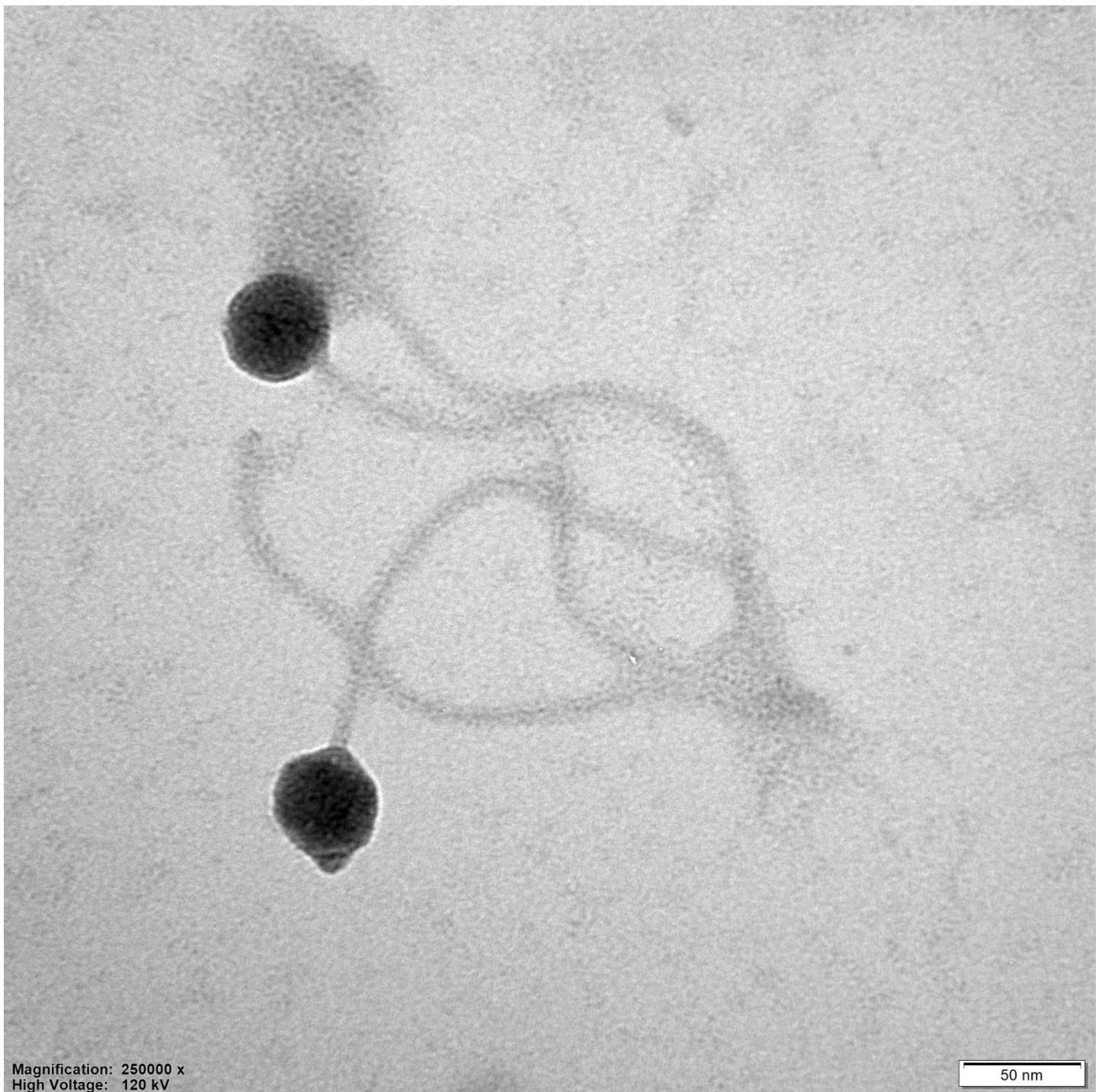


Fig. S1 phiCDKH01 morphology observed using a transmission electron microscope Tecnai Spirit BioTWIN microscope (FEI) at an acceleration voltage of 120 kV. Phage particles were stained with 2% uranyl acetate.

Table S1. ORFs locations, transcriptional orientations, sizes of encoded proteins, possible function, BLAST hit, alignment and E value

Locus tag	Coding region (bp)	Strand	Protein size (aa)	Predicted functional	BLAST hit [source organism] <sup>a</sup>	Accession number	Alignment (%ID)	E value
phiCDKH01_01	4-576	+	190	tail protein - partial	phage tail protein [ <i>Clostridioides difficile</i> ]	MBH6979172.1	190/636 (100%)	3e-125
phiCDKH01_02	611-2944	+	777	hypothetical protein	hypothetical protein [ <i>Clostridioides difficile</i> ]	WP_021411116.1	777/777 (100%)	0.0
phiCDKH01_03	2959-3285	+	108	hypothetical protein	hypothetical protein [ <i>Clostridioides difficile</i> ]	WP_021411117.1	108/108 (100%)	2e-69
phiCDKH01_04	3311-3463	+	50	hypothetical protein	hypothetical protein [ <i>Clostridioides difficile</i> ]	WP_021372475.1	50/50 (100%)	3e-25
phiCDKH01_05	3527-3793	+	88	holin	putative holin [ <i>Clostridium phage phiCD24-1</i> ]	CEK40241.1	88/88 (100%)	2e-63
phiCDKH01_06	3728-3504	+	268	N-acetylmuramoyl-L-alanine amidase	N-acetylmuramoyl-L-alanine amidase [ <i>Clostridioides difficile</i> ]	WP_021411118.1	268/268 (100%)	0.0
phiCDKH01_07	4644-4820	-	58	AbrB/MazE/SpoVT family DNA-binding domain protein	AbrB/MazE/SpoVT family DNA-binding domain-containing protein [ <i>Clostridioides difficile</i> ]	WP_021411119.1	58/58 (100%)	2e-33
phiCDKH01_08	4821-4988	-	55	hypothetical protein	hypothetical protein [ <i>Clostridioides difficile</i> ]	WP_021411120.1	55/55 (100%)	2e-28
phiCDKH01_09	5065-5193	-	42	membrane protein	putative membrane protein [ <i>Clostridioides difficile</i> P24]	EQJ50015.1	42/42 (69.05%)	1e-09
phiCDKH01_10	5517-5702	-	61	PemI protein	PemI [ <i>Clostridioides difficile</i> ]	WP_077725908	61/61	1e-34

phiCDKH01_11	5931-6059	-	42	hypothetical protein	hypothetical protein QUA_1656 [ <i>Clostridioides difficile</i> P49]	EQJ91741.1	<b>(96.72%)</b> 42/42 (100%)	3e-18
phiCDKH01_12	6577-7707	-	376	integrase	site-specific integrase [ <i>Clostridioides difficile</i> ]	WP_021411091.1	<b>376/376</b> (100%)	0.0
phiCDKH01_13	7948-8454	-	168	metallo-endopeptidase	MULTISPECIES: ImmA/IrrE family metallo-endopeptidase [ <i>Clostridiales</i> ]	WP_003429804.1	<b>168/168</b> (100%)	1e-114
phiCDKH01_14	8558-8980	-	140	TIR domain-containing protein	TIR domain-containing protein [ <i>Clostridioides difficile</i> ]	WP_021365371.1	<b>140/140</b> (100%)	2e-95
phiCDKH01_15	9044-9265	+	73	hypothetical protein	hypothetical protein [ <i>Clostridioides difficile</i> ]	WP_021365372.1	73/73 (100%)	5e-45
phiCDKH01_16	9279-9695	-	138	DUF4231 domain-containing protein	DUF4231 domain-containing protein [ <i>Clostridioides difficile</i> ]	WP_021365373.1	<b>138/138</b> (100%)	9e-90
phiCDKH01_17	9704-10249	-	181	transcriptional regulator	MULTISPECIES: XRE family transcriptional regulator [ <i>Clostridiales</i> ]	WP_003433494.1	<b>181/181</b> (100%)	6e-128
phiCDKH01_18	10294-10410	-	38	membrane protein	putative membrane protein [ <i>Clostridioides difficile</i> CD22]	EQE30741.1	<b>38/38</b> (100%)	3e-14

<b>phiCDKH01_19</b>	<b>10430-10645</b>	<b>+</b>	<b>71</b>	<b>transcriptional regulator</b>	<b>helix-turn-helix transcriptional regulator [<i>Clostridioides difficile</i>]</b>	<b>WP_021424794.1</b>	<b>71/71 (100%)</b>	<b>9e-43</b>
<b>phiCDKH01_20</b>	<b>10676-11494</b>	<b>+</b>	<b>272</b>	<b>antirepressor protein</b>	<b>phage antirepressor KilAC domain-containing protein [<i>Clostridioides difficile</i>]</b>	<b>WP_022583452.1</b>	<b>272/272 (100%)</b>	<b>0.0</b>
phiCDKH01_21	11543-11680	+	45	hypothetical protein	hypothetical protein [ <i>Clostridioides difficile</i> ]	WP_165479131.1	45/45 (100%)	1e-22
<b>phiCDKH01_22</b>	<b>11692-11859</b>	<b>-</b>	<b>55</b>	<b>DNA-binding protein</b>	<b>MULTISPECIES: Arc family DNA-binding protein [<i>Clostridiales</i>]</b>	<b>WP_021418310.1</b>	<b>55/55 (100%)</b>	<b>5e-30</b>
<b>phiCDKH01_23</b>	<b>11982-12158</b>	<b>+</b>	<b>58</b>	<b>HicB antitoxin</b>	<b>toxin-antitoxin system HicB family antitoxin [<i>Clostridioides difficile</i>]</b>	<b>WP_021424796.1</b>	<b>58/58 (100%)</b>	<b>1e-33</b>
<b>phiCDKH01_24</b>	<b>12151-12954</b>	<b>+</b>	<b>267</b>	<b>antirepressor protein</b>	<b>phage antirepressor KilAC domain-containing protein [<i>Clostridioides difficile</i>]</b>	<b>WP_103899391.1</b>	<b>267/267 (100%)</b>	<b>0.0</b>
phiCDK01_25	13002-13184	+	60	hypothetical protein	hypothetical protein [ <i>Clostridioides difficile</i> ]	WP_021365376.1	60/60 (100%)	5e-32
phiCDKH01_26	13142-13378	-	78	hypothetical protein	hypothetical protein [ <i>Clostridioides difficile</i> ]	WP_021424801.1	78/78 (100%)	6e-46
<b>phiCDKH01_27</b>	<b>13548-13940</b>	<b>+</b>	<b>130</b>	<b>helix-turn-helix domain-containing protein</b>	<b>helix-turn-helix domain-containing protein [<i>Clostridioides difficile</i>]</b>	<b>WP_021424802.1</b>	<b>130/130 (100%)</b>	<b>8e-86</b>

phiCDKH01_28	13927-14070	+	47	hypothetical protein	hypothetical protein [ <i>Clostridioides difficile</i> ]	WP_021390082.1	47/47 (100%)	1e-20
phiCDKH01_29	14237-14479	+	80	hypothetical protein	hypothetical protein [ <i>Clostridioides difficile</i> ]	WP_021365380.1	80/80 (100%)	5e-50
phiCDKH01_30	14558-14839	+	93	hypothetical protein	hypothetical protein [ <i>Clostridioides difficile</i> ]	WP_021365381.	93/93 (100%)	2e-60
<b>phiCDKH01_31</b>	<b>14826-15464</b>	<b>+</b>	<b>212</b>	<b>recombinase</b>	<b>recombinase [<i>Clostridioides difficile</i>]</b>	<b>PCD11566.1</b>	<b>212/213 (100%)</b>	<b>4e-151</b>
phiCDKH01_32	15474-16373	+	299	dnaD domain protein	dnaD domain protein [ <i>Clostridioides difficile</i> P49]	EQJ91696.1	299/299 (100%)	0.0
phiCDKH01_33	16393-16812	+	139	single-stranded DNA-binding protein	single-stranded DNA-binding protein [ <i>Clostridioides difficile</i> ]	WP_021411097.1	139/139 (100%)	3e-96
phiCDKH01_34	16893-16812	+	120	hypothetical protein	hypothetical protein [ <i>Clostridioides difficile</i> ]	WP_021424804.1	120/120 (100%)	9e-82
phiCDKH01_35	16842-17204	+	44	hypothetical protein	hypothetical protein CDIF29637_02647 [ <i>Clostridioides difficile</i> ]	AXU54373.1	44/44 (100%)	2e-21

phiCDKH01_36	17322-17441	+	39	hypothetical protein	hypothetical protein [ <i>Clostridioides difficile</i> ]	WP_021362348.1	39/39 (100%)	6e-19
phiCDKH01_37	17485-17598	+	37	hypothetical protein	hypothetical protein QU3_0844 [ <i>Clostridioides difficile</i> P42]	EQJ70600.1	37/37 (97.30%)	2e-15
phiCDKH01_38	17617-17814	+	65	hypothetical protein	hypothetical protein [ <i>Clostridioides difficile</i> ]	WP_021408087	65/65 (100%)	4e-39
<b>phiCDKH01_39</b>	<b>17841-18257</b>	<b>+</b>	<b>138</b>	<b>DUF1064 domain-containing protein</b>	<b><i>DUF1064 domain-containing protein</i> [<i>Clostridioides difficile</i>]</b>	<b>WP_021362351.1</b>	<b>138/138 (100%)</b>	<b>2e-89</b>
phiCDKH01_40	18276-18776	+	166	hypothetical protein	<i>hypothetical protein</i> [ <i>Clostridioides difficile</i> ]	WP_021362352.1	166/166 (100%)	1e-113
<b>phiCDKH01_41</b>	<b>19040-19171</b>	<b>+</b>	<b>43</b>	<b>membrane protein</b>	<b>putative membrane protein [<i>Clostridioides difficile</i> CD22]</b>	<b>EQE30759.1</b>	<b>43/37 (100%)</b>	<b>6e-15</b>
<b>phiCDKH01_42</b>	<b>19186-19362</b>	<b>+</b>	<b>58</b>	<b>PemI protein</b>	<b>putative PemI [<i>Clostridioides difficile</i>]</b>	<b>VIB43238.1</b>	<b>58/58 (98.28%)</b>	<b>9e-32</b>
<b>phiCDKH01_43</b>	<b>19788-20228</b>	<b>+</b>	<b>146</b>	<b>terminase small subunit</b>	<b>terminase small subunit [<i>Clostridioides difficile</i>]</b>	<b>WP_021365389.1</b>	<b>146/146 (100%)</b>	<b>1e-99</b>
<b>phiCDKH01_44</b>	<b>20221-21483</b>	<b>+</b>	<b>420</b>	<b>terminase large subunit</b>	<b>PBSX family phage terminase large subunit [<i>Clostridioides difficile</i>]</b>	<b>WP_021372000.1</b>	<b>420/420 (100%)</b>	<b>0.0</b>

phiCDKH01_45	21494-22978	+	494	portal protein	phage portal protein [ <i>Clostridioides difficile</i> ]	WP_021411100.1	494/494 (100%)	0.0
phiCDKH01_46	22978-24396	+	472	ADP-ribosyltransferase exoenzyme family protein	ADP-ribosyltransferase exoenzyme family protein [ <i>Clostridioides difficile</i> ]	WP_021411101.1	472/472 (100%)	0.0
phiCDKH01_47	24389-24565	+	58	hypothetical protein	hypothetical protein [ <i>Clostridioides difficile</i> ]	WP_021365393.1	58/58 (100%)	2e-33
phiCDKH01_48	24606-24836	+	76	ADP-ribosyltransferase exoenzyme family protein	ADP-ribosyltransferase exoenzyme family protein [ <i>Clostridioides difficile</i> P24]	EQJ49987.1	76/76 (100%)	1e-46
phiCDKH01_49	24837-25052	+	71	hypothetical protein	hypothetical protein [ <i>Clostridioides difficile</i> ]	WP_021365395.1	71/71 (100%)	3e-43
phiCDKH01_50	25082-25258	+	58	hypothetical protein	hypothetical protein [ <i>Clostridioides difficile</i> ]	WP_021365396.1	58/58 (100%)	6e-34
phiCDKH01_51	25394-26029	+	211	scaffolding protein	phage scaffolding protein [ <i>Clostridioides difficile</i> ]	WP_021372476.1	211/211 (100%)	8e-149
phiCDKH01_52	26047-27015	+	322	capsid protein	coat protein [ <i>Clostridioides difficile</i> ]	WP_131043024.1	322/322 (99.69%)	0.0
phiCDKH01_53	27041-27295	+	84	hypothetical protein	hypothetical protein [ <i>Clostridioides difficile</i> ]	WP_021365399.1	84/84 (100%)	2e-54
phiCDKH01_54	27292-27654	+	120	hypothetical protein	hypothetical protein [ <i>Clostridioides difficile</i> ]	WP_021411102.1	120/120 (100%)	1e-80

phiCDKH01_55	27656-28090	+	144	hypothetical protein	hypothetical protein [ <i>Clostridioides difficile</i> ]	WP_021411103.1	144/144 (100%)	1e-102
phiCDKH01_56	28098-28442	+	114	hypothetical protein	hypothetical protein [ <i>Clostridioides difficile</i> ]	WP_021365402.1	114/114 (100%)	2e-76
phiCDKH01_57	28448-29254	+	268	hypothetical protein	hypothetical protein [ <i>Clostridioides difficile</i> ]	WP_021372459.1	268/268 (100%)	0.0
phiCDKH01_58	29313-29789	+	158	hypothetical protein	hypothetical protein [ <i>Clostridioides difficile</i> ]	WP_021411104.1	158/158 (100%)	3e-108
phiCDKH01_59	29782-29979	+	65	hypothetical protein	conserved protein of unknown function [ <i>Clostridium phage phiCD24-1</i> ]	CEK40228.1	65/77 (100%)	1e-37
<b>phiCDKH01_60</b>	<b>30718-31251</b>	<b>+</b>	<b>177</b>	<b>lipoprotein</b>	<b>lipoprotein [<i>Clostridioides difficile</i>]</b>	<b>WP_021411109.1</b>	<b>177/177 (100%)</b>	<b>7e-127</b>
<b>phiCDKH01_61</b>	<b>31357-36885</b>	<b>+</b>	<b>1842</b>	<b>tail tape-measure protein</b>	<b>phage tail tape measure protein [<i>Clostridioides difficile</i>]</b>	<b>WP_021411110.1</b>	<b>1842/1842 (100%)</b>	<b>0.0</b>
<b>phiCDKH01_62</b>	<b>37323-38078</b>	<b>+</b>	<b>251</b>	<b>tail family protein</b>	<b>phage tail family protein [<i>Clostridioides difficile</i>]</b>	<b>WP_021411111.1</b>	<b>251/251 (100%)</b>	<b>0.0</b>



<b>phiCDKH01_63</b>	<b>38080-40839</b>	<b>+</b>	<b>919</b>	<b>tail protein</b>	<b>phage tail protein [Clostridioides difficile]</b>	<b>WP_021411112.1</b>	<b>919/919</b> <b>(100%)</b>	<b>0.0</b>
phiCDKH01_64	40826-41131	+	101	hypothetical protein	hypothetical protein [Clostridioides difficile]	WP_021411113.1	101/101 (100%)	7e-69
<b>phiCDKH01_65</b>	<b>41149-43368</b>	<b>+</b>	<b>739</b>	<b>pre-neck appendage-like protein</b>	<b>phage pre-neck appendage-like protein [Clostridioides difficile]</b>	<b>VIF60740.1</b>	<b>739/739</b> <b>(96.48%)</b>	<b>0.0</b>
<b>phiCDKH01_66</b>	<b>44414-45076</b>	<b>+</b>	<b>220</b>	<b>membrane protein</b>	<b>putative membrane protein [Clostridioides difficile P53]</b>	<b>ERM48055.1</b>	<b>220/169</b> <b>(96.45%)</b>	<b>1e-92</b>

<sup>a</sup>Only the most relevant hit is shown

Table S2. CRISPR spacer analysis

**A. phiCDKH01 CRISPR-spacer sequences**

spacer1: cattattaagtcaaagtgtgatgtaactgaacttaa  
spacer2: agtatatctaataataaatgagttcttagttaacc  
spacer3: tggacgagtatatctgaaggtgtagtagtgcttg  
spacer4: gaattgaatgataacaaaaaaaaagcatgtaatgaac  
spacer5: gctcataataaccataagcgcttttggttaaattaa

**B. Spacer match on published *C. difficile* phage genomes**

<u>phiCDKH01</u>	5 Hits	Length	Start	End	E value
<b>spacer1</b>	100%	36	30230	30265	1e-08
<b>spacer2</b>	100%	35	30296	30330	3e-08
<b>spacer3</b>	100%	35	30361	30395	3e-08
<b>spacer4</b>	100%	37	30426	30462	3e-09
<b>spacer5</b>	100%	37	30493	30529	3e-09
<u>phi027</u>	No Hit				
<u>phiC2</u>	No Hit				
<u>phiCD111</u>	No Hit				
<u>phiCD119</u>	No Hit				
<u>phiCD146</u>	No Hit				
<u>phiCD24-1</u>	No Hit				
<u>phiCD27</u>	No Hit				
<u>phiCD38-2</u>	No Hit				
<u>phiCD481-1</u>	No Hit				
<u>phiCD505</u>	No Hit				
<u>phiCD506</u>	No Hit				
<u>phiCD52</u>	No Hit				
<u>phiCD6356</u>	No Hit				
<u>phiCDHM1</u>	No Hit				
<u>phiCDHM11</u>	No Hit				
<u>phiCDHM19</u>	No Hit				
<u>phiE16</u>	No Hit				
<u>phiMMP01</u>	No Hit				
<u>phiMMP02</u>	No Hit				
<u>phiMMP03</u>	No Hit				
<u>phiMMP04</u>	No Hit				
<u>phiSemix9P1</u>	No Hit				
<u>phiCD630-01</u>	No Hit				
<u>phiCD630-02</u>	No Hit				
<u>phiLIBA2945</u>	No Hit				
<u>phiLIBA6276</u>	No Hit				

**C. Spacer match on published *C. difficile* strain genomes\***

<b>spacer2</b>	1 Hit
CP028529.1	100%
CP019858.1	100%
CP016104.1	100%
CP019469.1	100%
LK933005.1	100%
LK932396.1	100%
LK932486.1	100%
CP028361.1	100%
FN665652.1	100%

CP029154.1	100%
CP020424.2	100%
<b>spacer5</b>	1 Hit
CP028529.1	97.14%
CP019858.1	97.14%
CP019469.1	97.14%
LK933005.1	97.14%
LK932396.1	97.14%
LK932486.1	97.14%
CP049958.1	97.14%
CP028361.1	97.14%
FN665652.1	97.14%

\*removing phiCDKH01 containing strains