

ORF	location	Size (aa)	Most significant match (organism, E value); top match showing possible function	Domain/s ^a
1	43-456	137	putative terminase small subunit (<i>Staphylococcus aureus</i> , 1e-15)	IPR005335, terminase small subunit
2	559-1800	413	phage terminase, large subunit (<i>Clostridium perfringens</i> , 2e-120)	IPR006437, IPR006701 phage terminase, large subunit, PBSX
3	1811-3304	497	phage portal protein, SPP1 family (<i>Fusobacterium peridonticum</i> , 4e-79)	IPR006428, IPR021145, bacteriophage portal protein, SPP1 family
4	3189-3917	242	putative phage minor head protein (<i>Listeria seeligeri</i> , 1e-19)	IPR006528, phage putative head morphogenesis protein, SPP1 Gp7
5	3998-4615	205	hypothetical protein (<i>Bacillus</i> sp., 3e-26); putative ribonuclease phage-like protein (<i>Bacillus licheniformis</i> , 9e-24)	
6	4637-5539	300	putative major head protein (<i>Lacococcus lactis</i> , 2e-39)	
7	5622-5939	105	phage protein (<i>Bacillus cereus</i> , 4e-08)	IPR021146 bacteriophage QLRG family, putative DNA packaging
8	5940-6263	107	hypothetical protein (<i>Clostridium thermocellum</i> , 3.3)	
9	6244-6588	114	hypothetical protein (<i>Abiotrophia defectiva</i> 1.6)	
10	6572-6967	131	transketolase (<i>Moraxella catarrhalis</i> , 1)	
11	7010-7447	145	hypothetical protein (<i>Paenibacillus larvae</i> , 5e-25); phage major tail protein, TP901-1 family (<i>Bacillus cellulosilyticus</i> , 5e-20)	IPR011855 phage major tail protein TP901-1
12	7467-7820	117	hypothetical protein (<i>Haemophilus influenzae</i> , 3.0)	
13	7910-8122	70	DNA topoisomerase III (<i>Saccharophagus degradans</i> , 4.4)	
14	8126-10324	732	phage protein (<i>Clostridium ljungdahlii</i> , 2e-54); phage tape measure protein (<i>Finegoldia magna</i> , 2e-53)	IPR010090 phage tail tape measure protein TP901, core region
15	10325-11008	227	hypothetical protein (<i>Clostridium saccharolyticum</i> , 8e-14); phage tail protein (<i>Coprococcus catus</i> , 4e-13)	IPR008841 Siphovirus tail component
16	11010-12455	481	hypothetical protein (<i>Clostridium botulinum</i> , 4e-38); phage tail fibre protein (<i>Clostridium kluyveri</i> , 9e-13)	SSF47162 Apolipoprotein
17	12457-15135	892	probable transcriptional regulator-related protein (<i>Clostridium botulinum</i> , 1e-135)	
18	15149-15340	63	serine-type D-Ala-D-Ala carboxypeptidase (<i>Clostridium botulinum</i> , 0.024)	
19	15342-15494	50	hypothetical protein (<i>Clostridium botulinum</i> , 2e-7)	IPR010022 Phage uncharacterised protein XkdX
20	15507-17474	655	phage minor structural protein (<i>Listeria seeligeri</i> , 4e-30)	pfam06605 Prophage_tail
21	17511-17675	54	no hits	
22	17678-18472	264	N-acetylmuramoyl-L-alanine amidase (<i>Clostridium botulinum</i> , 6e-53)	IPR002508 cell wall hydrolase/autolysin, catalytic transmembrane regions
23	18482-18760	92	Imm (<i>Clostridium novyi</i> , 4e-04)	
24	18837-19193 complement	118	malonyl CoA-acyl carrier protein transacylase (<i>Veillonella</i> sp., 3.7)	

25	19398-19511	37	no hits	
26	19685-20296	203	hypothetical phage protein (<i>Clostridium</i> phage phiCTP1, 1e-13)	
27	20523-22646	707	hypothetical protein (Lachnospiraceae bacterium, 2e-163); DNA polymerase I (<i>Eubacterium eligens</i> , 1e-160)	IPR001098, IPR002298, DNA-directed DNA polymerase A; IPR002562 3'-5' exonuclease; IPR012337 Ribonuclease H-like IPR022765 Dna2/Cas4, DUF83
28	22714-23553	279	hypothetical protein (<i>Eubacterium eligens</i> , 7e-51); RecB (<i>Streptococcus</i> phage Dp-1, 1e-20)	
29	23553-24002	149	penicillin-binding protein-related factor A, putative recombinase (<i>Eubacterium rectale</i> , 5e-25)	IPR011335 restriction endonuclease, type II-like
30	23992-24222	76	no hits	
31	24238-24579	113	hypothetical protein (<i>Clostridium botulinum</i> , 2e-19)	
32	24653-24871	72	hypothetical protein (<i>Plasmodium chabaudi chabaudi</i> , 7)	
33	24864-25859	331	RecA protein (Lachnospiraceae bacterium, 3e-80)	IPR003593 ATPase, AAA+ type, core; IPR013765, IPR020587, IPR020588, IPR023400 DNA recombination and repair protein RecA
34	25861-26040	59	no hits	
35	26030-26353	107	hypothetical protein (<i>Streptococcus</i> phage Dp-1, 1e-18)	
36	26368-26598	76	esterase D/formylglutathione hydrolase (<i>Rattus norvegicus</i> , 5.2)	
37	26600-26812	70	aldehyde dehydrogenase (<i>Burkholderia ambifaria</i> , 5)	
38	26812-26991	59	no hits	
39	27102-27614	170	hypothetical protein (<i>Ralstonia pickettii</i> , 1.1)	
40	27616-27738	40	no hits	
41	27740-27868	42	no hits	
42	27880-29583	567	PhoH family protein (<i>Verrucosipora maris</i> , 4e-65); replicative DNA helicase (<i>Burkholderia cenocepacia</i> , 3e-60)	IPR004042, IPR006142 intein DOD homing endonuclease; IPR007868 Hom-end-associated Hint; IPR013317 chromosomal replication control, initiator (DnaA)/regulator (Hda) IPR007694 DNA helicase, DnaB-like, C-terminal
43	29576-30823	415	replicative DNA helicase (<i>Eubacterium eligens</i> , 2e-78)	
44	30861-31031	56	hypothetical protein (<i>Clostridium botulinum</i> , 1e-15)	
45	31032-31232	66	aromatic amino acid aminotransferase (<i>Ruegeria</i> sp., 8.4)	
46	31244-33331	695	hypothetical protein (<i>Clostridium hylemonae</i> , 3e-141); anaerobic ribulose triphosphate reductase (<i>Clostridium</i> sp., 2e-140)	IPR005144 ATP-cone; IPR012833 ribonucleoside-triphosphate reductase, anaerobic
47	33507-33854	115	NAD-dependent DNA ligase <i>Streptococcus</i> phage DP-1, 2e-14	COG0272 NAD-dependent DNA ligase
48	33867-34244	125	aldehyde dehydrogenase (NAD(+)) (<i>Corynebacterium striatum</i> , 0.43)	
49	34244-34723	159	anaerobic NTP reductase, small subunit	IPR012837 ribonucleoside-

			(<i>Enterobacteria</i> phage Felix 01, 1e-30)	triphosphate reductase activity, anaerobic
50	34711-35601	296	DNA primase (<i>Eubacterium eligens</i> , 2e-33)	IPR002694 zinc finger, CHC2-type
51	35598-35789	63	GG21719 (<i>Drosophila erecta</i> , 6)	
52	35789-36127	112	nuclear pore complex subunit <i>Saccharomyces cerevisiae</i> 0.43	
53	36129-36620	163	hypothetical phage protein (<i>Clostridium</i> phage phiCTP1, 5e-36); putative deoxyuridine 5'-triphosphate nucleotidohydrolase (<i>Bacillus amyloliquefaciens</i> , 1e-29)	
54	36731-37669	312	group I glycosyl transferase (<i>Chloroflexus aggregans</i> , 0.45)	
55	37669-38292	207	hypothetical protein (<i>Clostridium botulinum</i> , 3e-11)	
56	38285-38581	98	RNA polymerase, sigma-factor (<i>Algoriphagus</i> sp., 0.32)	
57	38548-38736	62	UvrABC system protein (<i>Streptomyces clavuligerus</i> , 1.1)	
58	38754-39509	251	hypothetical protein (<i>Staphylococcus</i> phage K, 6e-18); nucleoside 2-deoxyribosyltransferase family (<i>Clostridium botulinum</i> , 1e-08)	pfam05014, nucleoside 2-deoxyribosyltransferase
59	39509-39772	87	hypothetical protein (<i>Meyerozyma guilliermondii</i> , 3.4)	
60	39765-40187	140	n/a (<i>Ectocarpus siliculosus</i> , 2)	
61	40187-40438	83	hypothetical protein (<i>Clostridium botulinum</i> , 2e-6)	pfam04193 PQ loop repeat
62	40489-40854	121	intron-associated endonuclease (<i>Enterobacteria</i> phage T4, 2e-6)	IPR000305 excinuclease ABC, C subunit
63	40900-41133	77	sugar diacid utilization factor (<i>Lactobacillus farciminis</i> , 6)	
64	41272-41421	49	no hits	
65	41448-41720	90	histidine sensor kinase protein (<i>Aquifex aeolicus</i> , 6.6)	
66	41952-42125	57	no hits	
67	42188-42358	56	no hits	
68	42402-42617	71	hypothetical protein (<i>Paramecium tetraurelia</i> , 1.5)	
69	42604-42750	48	no hits	
70	42851-43075	74	no hits	
71	43950-44078 complement	42	no hits	
72	44090-44200 complement	36	no hits	
73	44345-44698 complement	117	possible glutamate synthase (NADPH) (<i>Oribacterium sinus</i> , 8.9)	
74	44726-44821 complement	31	no hits	
75	44858-45409	185	hypothetical protein (<i>Shewanella</i>	

76	complement 45409-45747	112	<i>amazonensis</i> , 5.6) phage protein (<i>Clostridium papyrosolvens</i> , 5e-09)	IPR010359, DUF955
77	complement 45820-45990	56	no hits	
78	complement 45993-46148	51	no hits	
79	complement 46138-46296	52	no hits	
80	complement 46311-46469	52	no hits	
81	complement 46523-46762	79	hypothetical protein (<i>Nectria haematococca</i> , 7.7)	
82	complement 46783-47124	113	transposase (<i>Burkholderia rhizoxinica</i> , 0.74)	pfam09339 IciR helix-turn-helix domain

^a Domain or family assignments from InterProScan (IPR), pfam or Superfamily (SSF).

TABLE S1. ORF analysis of the Φ 8074-B1 genome.