

Table S1. Features of PR1 predicted CDSs. For each CDS, the transcription strand, start and stop position is given. At protein level, the corresponding gene product size, molecular weight and pI as well as the homolog, predicted function, homology values and motifs are shown.

ORF (+/-)	Start (bp)	Stop (bp)	Size (aa)	MW (Kda)	pI	Best species hit ^a	Putative function	E-value (% identify)	Motifs	Homologous T5 ORF/T5p ^b
1/-	63	302	79	8.54	4.27	-	-	-	-	-
2/-	399	1,115	238	27.37	4.85	<i>Yersinia</i> phage phiR201*	deoxynucleoside-5'-monophosphatase (<i>dmp</i>)	7E-71 (0.47)	-	001/001
3/-	1,201	1,665	154	16.83	3.99	-	-	-	-	-
4/-	1,802	3,565	587	65.17	6.58	<i>Escherichia</i> phage DT57C*	DNA transfer protein A1	2E-164 (0.47)	Metallophos_2: (PF12850)	004/004
5/-	3,641	4,105	154	16.93	4.54	<i>Escherichia</i> T5*	DNA transfer protein A2	4E-8 (0.28)	-	007/006
6/-	4,376	4,666	96	11.04	9.17	-	-	-	TPR	-
7/-	4,772	5,113	113	13.00	9.10	-	-	-	-	-
8/+	5,468	6,547	359	27.28	4.44	<i>Pectobacterium</i> phage My1 *	hypothetical protein	7E-19 (0.39)	-	011_012/010
9/+	6,596	6,997	133	15.03	6.49	-	-	-	-	-
10/+	7,181	7,297	38	4.26	3.17	-	-	-	1 TMD	-
11/+	7,333	7,476	47	5.36	7.52	-	-	-	1 TMD	-
12/+	7,548	7,820	90	10.65	4.19	<i>Vibrio</i> phage 11895-B1	hypothetical protein	9E-05 (0.74)	-	-
13/+	7,907	8,335	142	15.72	4.19	-	-	-	-	-
14/+	8,579	8,848	89	9.63	8.47	-	-	-	-	-
15/+	8,910	9,041	43	5.20	9.89	-	-	-	-	-
16/+	9,105	9,335	76	8.42	4.13	-	-	-	-	-
17/+	10,551	10,694	47	5.56	4.56	-	-	-	-	-
18/-	11,044	11,940	298	33.54	7.64	-	-	-	-	-
19/-	11,986	12,687	233	26.54	4.73	-	-	-	-	-
20/-	12,885	13,181	98	11.61	5.91	-	-	-	-	-
21/-	13,295	13,513	72	7.90	0.41	-	-	-	1 TMD; 1 SP	-
22/-	13,510	13,782	90	9.06	4.31	-	-	-	-	-
22A/-	13,792	13,884	30	3.32	3.75	-	-	-	1 TMD	-
23/-	13,887	14,516	209	23.84	5.36	<i>Bacteroides</i> sp. CAG:875	guanosine-3'5'-bis(Diphosphate) 3'-pyrophosphohydrolase	1E-12 (0.34)	HD_4: (PF13328)	-
24/-	14,513	14,821	102	12.24	7.67	-	-	-	-	-
25/-	14,818	15,165	115	13.23	4.79	-	-	-	1 TMD	-
26/-	15,177	15,509	110	12.55	0.26	-	-	-	-	-
27/-	15,516	15,842	108	11.98	3.97	-	-	-	-	-
28/-	15,832	16,023	63	7.10	4.41	-	-	-	-	-
29/-	16,005	16,361	118	14.01	10.0	-	-	-	2 TMD	-
30/-	16,327	16,749	140	15.75	4.76	-	-	-	Acetyltransf_7: (PF13508)	-

31/-	16,746	17,039	97	11.12	5.13	-	-	-	-	-
32/-	17,027	17,206	59	6.76	7.17	-	-	-	1 TMD; 1 SP	-
33/-	17,314	17,559	81	8.90	10.15	-	-	-	2 TMD; 1 SP	-
34/-	17,560	18,012	150	6.94	8.26	<i>Pectobacterium</i> phage My1*	hypothetical protein	8E-10 (0.45)	-	022/022
35/-	17,969	18,424	151	17.31	4.31	<i>Pectobacterium</i> phage My1*	hypothetical protein	2E-18 (0.36)	-	024/025
36/-	18,424	18,750	108	12.17	5.73	-	-	-	-	-
37/-	18,754	19,122	122	13.80	8.43	-	-	-	1 TMD	-
38/-	19,122	19,304	60	7.02	10.62	-	-	-	1 SP	-
39/-	19,306	19,599	97	11.09	10.07	-	-	-	-	-
40/-	19,674	20,042	122	14.15	4.67	-	-	-	-	-
41/-	20,045	20,848	267	30.32	5.12	<i>Salmonella</i> phage SPC35*	serine/threonine protein phosphatase	1E-44 (0.37)	Metallophos_2: (PF12850)	031/034
42/-	20,880	21,428	182	21.10	8.09	<i>Escherichia</i> phage DT57C*	phosphatase	9E-36 (0.45)	Metallophos_2: (PF12850)	030/033
43/-	21,431	21,583	50	5.81	3.50	-	-	-	Gly_radical: (PF01228)	-
44/-	21,586	22,020	144	16.20	9.14	<i>Vibrio</i> phage phi 3*	phage protein	0.00006 (0.31)	-	034/037
45/-	22,098	22,523	141	15.82	9.34	<i>Klebsiella</i> phage KP27	endolysin	3E-34 (0.48)	VanY: (PF02557)	036/039
46/-	22,523	23,191	222	25.45	5.10	<i>Pectobacterium</i> phage My1*	holin	3E-69 (0.48)	Phage_holin_T: (PF11031); 1 TMD	037/040
47/-	23,358	24,062	234	25.96	4.34	<i>Pectobacterium</i> phage My1*	Clp protease	2E-42 (0.46)	CLP_protease: (PF00574)	038/041
48/-	24,074	24,778	234	26.21	4.99	<i>Escherichia</i> phage AKFV33*	Deoxynucleoside monophosphate kinase	1E-24 (0.36)	P-mevalo_kinase: (PF04275)	039/042
49/-	24,768	25,091	107	12.27	8.29	<i>Salmonella</i> phage STP4-a	outer membrane lipoprotein, o-spanin	1E-13 (0.28)	SpII cleavage site; Pilus_CpaD: (PF09476); 1 SP	040/043
50/-	25,043	25,495	113	12.64	7.25	<i>Escherichia</i> phage T5*	cytoplasmic membrane protein, i-spanin	7E-13 (0.36)	-	041/044
51/-	25,479	25,880	133	15.11	4.75	-	-	-	-	-
52/-	25,870	26,115	81	8.86	11.21	-	-	-	1 TMD	-
53/-	26,175	26,513	112	13.01	8.41	-	-	-	-	-
54/-	26,696	27,187	163	17.85	4.10	<i>Serratia</i>	radical cofactor pyruvate-formate lyase	4E-36 (0.56)	Gly_radical: (PF01228); 1 TMD	054/055
55/-	27,198	27,452	84	8.96	4.36	-	-	-	1 SP	-
56/-	27,454	27,813	119	13.52	3.87	-	-	-	-	-
57/-	27,816	28,367	183	20.12	7.51	-	-	-	AAA_14: (PF13173)	-
58/-	28,751	28,924	57	6.67	9.14	-	-	-	-	-
59/-	28,933	29,121	62	7.14	9.51	-	-	-	1 SP	-
60/-	29,259	29,372	37	4.33	5.41	-	-	-	-	-
61/-	29,374	29,895	173	19.13	9.76	<i>Escherichia</i> phage DT57C*	DNA primase	7E-25 (0.41)	-	071
62/-	30,074	30,238	54	6.36	4.23	-	-	-	-	-

63/-	30,404	30,694	96	11.34	4.29	-	-	-	-	-	-
64/-	30,691	30,909	72	8.25	7.12	-	-	-	-	-	-
65/-	31,002	31,553	183	21.73	4.22	-	-	-	-	-	-
66/-	31,935	32,066	43	5.08	6.79	-	-	-	-	-	-
67/-	32,150	32,317	55	5.57	10.13	-	-	-	2 TMD; 1 SP	-	-
68/-	32,415	32,726	103	12.13	7.75	-	-	-	-	-	-
69/-	33,114	33,503	129	14.42	4.56	-	-	-	-	-	-
70/-	33,503	33,688	61	6.92	5.74	-	-	-	-	-	-
71/-	33,757	33,966	69	8.06	10.31	-	-	-	-	-	-
72/-	34,051	34,374	107	12.71	8.78	-	-	-	-	-	-
73/-	34,374	34,730	118	13.62	5.14	<i>Yersinia</i> phage phiR1-37	Hypothetical protein	4E-15 (0.39)	Response_reg: (PF00072)	-	-
74/-	34,717	34,839	40	4.48	9.26	-	-	-	1 TMD	-	-
75/-	34,836	35,018	60	6.75	4.88	-	-	-	-	-	-
76/-	35,906	36,082	58	6.64	7.08	-	-	-	DnaJ_CXXCXGXG: (PF00684)	-	-
77/-	36,254	36,400	48	5.72	10.01	-	-	-	-	-	-
78/-	37,009	37,230	73	8.51	3.92	-	-	-	-	-	-
79/-	37,312	37,461	49	5.53	6.62	-	-	-	-	-	-
80/-	37,561	37,752	63	7.17	5.10	-	-	-	-	-	-
81/-	37,797	38,270	157	18.25	3.97	-	-	-	TPR	-	-
82/-	38,805	39,374	189	21.54	8.63	<i>Erwinia</i> phage Ea35-70	hypothetical protein	9E-63 (0.55)	LemA: (PF04011); 1 TMD; 1 SP	-	-
83/-	39,374	40,420	348	39.69	6.51	<i>Erwinia</i> phage Ea35-70	hypothetical protein	7E-82 (0.4)	3 TMD; 1 SP	-	-
84/-	40,540	40,920	126	13.60	3.67	-	-	-	-	-	-
85/-	40,996	41,361	121	13.75	3.82	-	-	-	-	-	-
86/-	41,472	41,720	82	9.03	3.59	-	-	-	-	-	-
87/-	41,799	42,197	132	14.69	3.97	<i>Microgenomates</i> bacterium	hypothetical protein	2E-28 (0.46)	-	-	-
88/+	42,601	42,864	87	10.15	10.50	-	-	-	1 TMD	-	-
89/-	42,938	43,222	94	10.74	6.26	-	-	-	-	-	-
90/-	43,230	43,664	144	16.26	5.31	<i>Salmonella</i> phage Shivani*	hypothetical protein	8E-11 (0.36)	GatB_Yqey: (PF02637)	-	076
91/-	43,666	44,103	145	16.53	9.54	<i>Pectobacterium</i> phage My1*	cell wall hydrolase	8E-21 (0.39)	SleB: (COG3773); 1 SP	-	078/081
92/-	44,103	44,480	125	13.83	10.22	-	-	-	-	-	-
93/-	44,492	45,220	242	27.45	5.98	<i>Escherichia</i> phage slur09*	hypothetical protein	0.00003 (0.25)	RNase_T: (PF00929); 1 TMD	-	082/085
93A/-	45,162	45,455	97	11.33	4.88	<i>Pectobacterium</i> phage My1*	hypothetical protein	0.00009 (0.47)	-	-	083/086
94/-	45,406	45,783	125	14.55	11.19	<i>Escherichia</i> phage T5*	hypothetical protein	0.000007	-	-	084/087
95/-	45,871	46,149	92	10.62	9.36	<i>Pectobacterium</i> phage My1*	hypothetical protein	3E-19 (0.58)	-	-	085/088

96/-	46,217	46,714	165	18.61	5.69	<i>Cupriavidus</i> sp. BIS7	ribonuclease HI	2E-28 (0.44)	RNase_H: (PF00075)	086/089
97/-	46,714	47,010	98	10.83	4.02	-	-	-	-	-
98/-	47,115	47,960	281	31.74	6.67	<i>Idiomarina</i> <i>zobellii</i>	thymidylate synthase	1E-151 (0.73)	Thymidylat_synt: (PF00303)	087/090
99/-	47,957	48,484	175	19.26	4.37	<i>Salmonella</i> phage Stitch*	dihydrofolate reductase, phage-associated	5E-12 (0.31)	DHFR_1: (PF00186)	088/091
100/-	48,484	49,587	367	42.28	4.49	<i>Salmonella</i> phage Shivani*	ribonucleotide-diphosphate reductase class Ia (aerobic), beta subunit	9E-126 (0.52)	-	089/092
101/-	49,661	52,045	794	89.20	6.14	<i>Escherichia</i> phage vB_EcoS_FFH1*	ribonucleotide-diphosphate reductase class Ia (aerobic), alpha subunit	0.0 (0.58)	Ribonuc_red_IgN: (PF00317)	091/094
102/-	52,173	52,892	239	26.96	8.91	<i>Yersinia</i> phage phiR201*	phosphate starvation-inducible protein <i>PhoH</i> , predicted ATPase	2E-101 (0.6)	PhoH: PhoH-like prote: (PF02562)	092/096
103/+	53,186	55,057	623	70.09	5.86	<i>Salmonella</i> phage Stitch*	ribonucleotide reductase of class III (anaerobic), large subunit	0.0 (0.73)	NRDD: (PF13597)	093/097
104/+	55,123	55,476	117	12.65	4.30	-	-	-	-	-
105/+	55,469	55,681	70	7.82	4.10	<i>Salmonella</i> phage FelixO1*	hypothetical protein	1E-15 (0.54)	-	059/060
106/+	55,665	55,961	98	11.02	5.11	-	-	-	-	-
107/+	55,961	56,260	99	11.50	4.39	-	-	-	-	-
108/+	56,976	59,756	926	105.36	5.70	<i>Salmonella</i> phage SPC35*	replication origin binding protein	0.0 (0.65)	Herpes_ori_bp: (PF02399)	100/106
109/+	59,749	59,988	79	9.39	10.56	-	-	-	-	-
110/+	60,046	60,765	239	27.28	5.54	<i>Yersinia</i> phage phiR201*	hypothetical protein	8E-62 (0.48)	HTH_Tnp_IS630: (PF01710)	102/109
111/+	60,765	61,019	84	9.78	5.15	-	-	-	-	-
112/+	61,096	61,494	132	14.21	5.36	<i>Escherichia</i> phage T5*	hypothetical protein	6E-26 (0.46)	-	104/111
113/+	61,507	61,836	109	12.56	9.49	<i>Escherichia</i> phage DT57C*	hypothetical protein	3E-21 (0.48)	-	105/112
114/+	61,877	62,170	97	11.11	4.76	<i>Salmonella</i> phage Stitch*	transcriptional regulator protein	5E-26 (0.52)	PC4: Transcriptional (PF02229)	106/113
115/+	62,269	62,718	149	17.10	7.51	-	-	-	DNA_ligase_aden: (PF01653)	-
116/+	62,709	63,710	333	37.12	4.54	<i>Yersinia</i> phage phiR201*	DNA ligase	3E-114 (0.55)	DNA_ligase_OB (PF03120)	107/114
117/+	63,903	64,649	248	27.18	8.66	<i>Escherichia</i> phage vB_EcoS_FFH1*	DNA ligase subunit B	4E-78 (0.55)	Lig NAD-dependent DNA (COG0272)	108/115
118/+	64,656	65,444	262	28.55	6.33	<i>Pectobacterium</i> phage My1*	hypothetical protein	3E-89 (0.58)	Myb_DNA-bind_6: (PF13921)	109/116
119/+	65,546	67,075	505	55.91	4.97	<i>Yersinia</i> phage phiR201*	DNA helicase	8E-146 (0.47)	DnaB Replicative DNA helicase (COG0305)	110/117
120/+	67,072	67,950	292	32.63	6.05	<i>Yersinia</i> phage phiR201*	DNA replication primase	4E-121 (0.59)	DnaG DNA primase (COG0358)	112/119
121/+	68,023	69,831	875	69.08	5.71	<i>Escherichia</i> phage vB_EcoS_FFH1*	DNA polymerase exon	0.0 (0.68)	- SSL2 Superfamily II (COG1061)	113/120
122/+	70,092	70,877	261	30.22	8.42	<i>Domibacillus</i> <i>tundrae</i>	hypothetical protein	8E-23 (0.35)	LAGLIDADG_WhiA: (PF14527)	-
123/+	70,937	71,761	274	30.57	4.52	<i>Yersinia</i> phage phiR201*	DNA polymerase exon	7E-105 (0.64)	DNA_pol_A: (PF00476)	113/120
124/+	71,764	72,258	164	19.67	10.15	<i>Salmonella</i> phage Stitch*	hypothetical protein	2E-41 (0.47)	-	114/121

125/+	72,255	73,631	458	52.00	6.68	<i>Escherichia</i> phage DT57C*	helicase	9E-162 (0.55)	SSL2 Superfamily II DNA or RNA helicase (COG1061)	115/122
126/+	73,663	73,917	84	9.89	4.24	-	-	-	-	-
127/+	73,985	74,359	124	14.52	4.71	<i>Escherichia</i> phage AKFV33*	hypothetical protein	6E-32 (0.51)	-	116/123
128/+	74,346	75,164	272	30.50	4.90	<i>Escherichia</i> phage DT57C*	hypothetical protein	2E-39 (0.36)	-	117/124
129/+	75,217	76,194	325	37.19	6.05	<i>Pectobacterium</i> phage My1*	exonuclease subunit 1	1E-103 (0.5)	-	118/125
130/+	76,187	78,040	617	68.95	4.66	<i>Escherichia</i> phage slur09*	exonuclease subunit 2	1E-168 (0.44)	-	119/126
131/+	78,040	78,522	160	18.37	6.52	<i>Escherichia</i> phage vB_EcoS_FFH1*	hypothetical protein	8E-71 (0.62)	Hjc: (PF01870)	120/127
132/+	78,522	79,403	293	33.68	5.22	<i>Escherichia</i> phage slur09*	flap endonuclease	1E-88 (0.49)	5_3_exonuc_N: (PF02739)	121/128
133/+	79,396	79,860	154	16.89	5.92	<i>Vibrio</i> phage phi 3*	deoxyUTP pyrophosphatase	3E-38 (0.44)	dUTPase: (PF00692)	122/129
134/-	79,901	82,180	759	81.40	6.82	<i>Providencia</i> sneebia DSM 19967	tail fiber protein (L-shapped fibers)	1E-71 (0.47)	phage_tail_N: (PF08400)	124/131
135/-	82,180	82,593	137	15.12	4.92	<i>Salmonella</i> phage Stitch*	tail protein (L-fiber(tail tip collar))	2E-27 (0.42)	-	125/132
136/-	82,598	85,714	1038	110.76	8.18	<i>Salmonella</i> phage SPC35*	tail protein (central straight fiber)	1E-173 (0.45)	-	126/133
137/-	85,707	88,550	947	106.77	5.80	<i>Escherichia</i> phage AKFV33*	tail protein	0.0 (0.57)	Phage-tail_3: (PF13550)	127/134
138/-	88,551	89,159	202	22.50	5.80	<i>Salmonella</i> phage SPC35*	distal tail protein	1E-72 (0.57)	P22_CoatProtein: (PF11651)	128/135
139/-	89,276	92,989	1237	132.99	5.06	<i>Salmonella</i> phage Stitch*	tape measure protein	0.0 (0.36)	-	129/136
140/-	93,069	93,398	109	12.14	4.82	<i>Pectobacterium</i> phage My1*	hypothetical protein	2E-17 (0.39)	-	131/138
141/-	93,484	93,894	136	15.47	4.71	<i>Pectobacterium</i> phage My1*	hypothetical protein	8E-34 (0.48)	-	132/139
142/-	93,902	94,795	270	30.58	5.72	<i>Escherichia</i> phage AKFV33*	minor tail protein	6E-56 (0.42)	1 TMD	133/140
143/-	94,799	95,944	381	41.72	4.27	<i>Salmonella</i> phage Stitch*	major tail protein (tail tube protein)	9E-165 (0.6)	-	134/141
144/-	95,970	96,455	161	18.45	5.04	<i>Escherichia</i> phage DT57C*	tail tube terminator protein	2E-36 (0.41)	Phage_tail_U: (PF06141)	135/142
145/-	96,455	97,216	253	28.01	9.97	<i>Escherichia</i> phage EPS7*	hypothetical protein	3E-88 (0.55)	-	136/143
146/-	97,216	97,734	172	19.80	5.26	<i>Yersinia</i> phage phiR201*	hypothetical protein	9E-17 (0.3)	-	137/144
147/-	97,800	99,227	475	52.09	5.04	<i>Escherichia</i> phage slur09*	major capsid protein	8E-153 (0.56)	Phage_connect_1: (PF05135)	138/145
148/-	99,234	99,812	192	21.51	5.06	<i>Escherichia</i> phage vB_EcoS_FFH1*	prohead protease	8E-74 (0.62)	Phage_capsid: (PF05065)	139/146
149/-	99,809	101,020	403	44.10	4.38	-	-	-	Peptidase_S78: (PF04586)	-
150/-	101,017	102,258	413	45.95	5.53	<i>Yersinia</i> phage phiR201*	portal protein	0.0 (0.63)	Phage_portal: (PF04860)	141/148
151/-	102,258	102,761	167	19.54	6.96	<i>Escherichia</i> phage slur09*	putative nicking endonuclease, (<i>SciA</i>)	1E-45 (0.68)	HNH endonuclease: (PF01844)	142/149
152/-	102,941	104,263	440	50.57	5.35	<i>Escherichia</i> phage DT57C*	terminase large subunit	0.0 (0.71)	Terminase_3: (PF04466)	144/151

153/-	104,263	104,736	157	17.73	5.30	<i>Vibrio</i> phage phi 3*	terminase, small subunit, single-chain interruption protein B (<i>SciB</i>)	2E-36 (0.46)	-	145/152
154/-	104,746	106,611	621	67.73	9.40	-	probable receptor binding protein (RBP)	-	-	-
155/+	106,776	107,294	172	19.95	5.67	-	-	-	-	-

^aNames of phages of *T5virus* genus are marked by an asterisk. The assignment of phages to the *T5virus* genus is based on Davison (2015) (1) or on taxonomic data in NCBI nucleotide database.

^bThe assignment of T5-encoded homologs to predicted PR1 proteins was based on results of BLASTP analysis against viral protein database. In the case of low homologies, putative T5-encoded counterparts of PR1 proteins (marked with an asterisk) were identified if they appeared as top hits (E value < -6) in a search against the translated DNA sequence of PR1 and a given T5 protein sequence as a query and if the relative location of their genes in the PR1 genome corresponded to that of a query gene in the genome of T5. The assignment of numbers to T5 ORFs/proteins is as in Wang et al. 2005/Zivanovic et al., 2014 (2, 3). Functional assignments of PR1 genes encoding homologs of T5 morphogenetic proteins are as in Zivanovic et al., 2014 (2). Acronyms: TMD – transmembrane domain; SP – Signal peptide; TPR P – tetra tri copeptide repeat motif.

Table S2. tRNA genes and codon usage of PR1. The anti-codon for which each tRNA (anti-codon) matches is highlighted in bold.

tRNA	Amino acid	Phage codon and frequency (%)						Anti-codon	Size (bp)	GC %	Start	End	
1	Met	ATG						CAT	73	53.2	28,388	28,460	
2	Gln	GAG* GAA						TTG	73	54.8	28,591	28,663	
		818 1,547											
3	Trp	TGG						CCA	71	54.9	30,930	31,000	
		416											
4	Leu	TTG	TTA	CTG	CTA*	CTT	CTC	TAG	74	50.0	31,862	31,935	
		295	1,069	272	780	461	90						
5	Leu	TTG	TTA	CTG	CTA*	CTT	CTC	TAA	74	52.7	32,075	32,148	
		295	1,069	272	780	461	90						
6	Cys	TGT* TGC							GCA	71	46.5	32,337	32,407
		210 107											
7	Asn	AAT	AAC*				GTT	82	52.4	35,041	35,122		
		918 717											
8	Ile	ATA* ATT	ATC				GAT	74	51.3	35,132	35,205		
		537 1,014	626										
9	Glu	GAG* GAA							TTC	75	49.3	35,511	35,585
		818 1,547											
10	Tyr	TAT TAC*							GTA	82	54.9	35,669	35,750
		709 514											
11	Arg	AGG* AGA	CGG	CGA	CGT	CGC	ACG	73	49.3	35,843	35,915		
		133 447	32 145	564	157								
12	Thr	ACG ACA*	ACT	ACC				TGT	73	53.4	36,101	36,173	
		185 655	879 271										
13	Val	GTG GTA*	GTT	GTC				TAC	71	49.2	36,182	36,252	
		283 913	901 157										
14	His	CAT CAC*							GTG	75	56.0	36,481	36,555
		398 225											
15	Phe	TTT TTC*							GAA	72	50.0	36,699	36,770
		800 559											
16	Asp	GAT GAC*							GTC	74	55.4	36,779	36,852
		1,398 704											
17	Ala	GCG* GCA	GCT	GCC				TGC	73	50.7	36,862	36,934	
		264 969	1,070 243										
18	Pro	CCG CCA	CCT	CCC*				TGG	74	51.3	36,945	37,018	
		122 525	474 165										
19	Ser	AGT AGC	TCC*	TCT	TCC*	TCA	GCT	91	50.5	37,480	37,570		
		445 268	113 714	208 463									
20	Lys	AAG* AAA							TTT	73	49.3	38,291	38,363
		858 1,363											
21	Met	ATG						CAT	73	56.2	40,459	40,531	
		100											
22	Arg	AGG* AGA	CGG	CGA	CGT	CGC	TCT	72	47.2	41,393	41,464		
		133 447	32 145	564 157									

*less frequent *Providencia rettgeri* codons consulted at the Codon Usage Database (Kazusa DNA Research Institute, Japan)

Table S3. Location and nucleotide sequence of nick regions in PR1 DNA.

Nick name	Sequence in nicked 3'-5' strand (5'-3') ^a	Nick location in the genome			Fraction of phage DNA with a nick ^e
		position ^b	% genome length ^c	relative location ^d	
nick 1 (major)	A/ <u>GCGCAGAG</u>	9699/	8.2	NC, CDS16/CDS17*	45%
nick 2	G/ <u>GCGCAGGA</u>	26607/	22.4	NC, CDS54/CDS55	31%
nick 3	G/ <u>GCGCAGGA</u>	37780/	31.9	NC, CDS80/CDS81	24%
nick 4	G/ <u>GCGCAGGA</u>	52111/	44.0	NC, CDS101/CDS102	39%
nick 5	G/ <u>GCGCACGA</u>	67548/	57.0	CDS120 (DNA primase)	26%
nick 6	G/ <u>GCGCATAA</u>	88044/	74.3	CDS137	41%
nick 7	G/ <u>GCGCGATA</u>	97564/	82.3	CDS146	21%
nick 8	G/ <u>GCGCATGG</u>	103262/	87.1	CDS152 (terminase)	22%
nick 9 (major)	A/ <u>GCGCAGAG</u>	117775/	99.4	NC, CDS16/CDS17*	45%
consensus ^f	R/ <u>GCGC</u> RNDR				

^a Sequences of nick regions represent the sequences in 3'-5' strand where the nicks are located. Nucleotide residues conserved in all nick regions are underlined.

^b Nick position is shown relative to the corresponding position in 5'-3' strand

^c The % of genome length where a nick is located is calculated from the beginning of left LTR

^d NC stands for non-coding region, nicks in the redundant PR1 virion DNA regions (LTRs) are indicated by an asterisk

^e The fraction of phage DNA with a nick was estimated based on the number of sequence reads that start from the nick site as compared to the number of reads that start from the nick site plus the number of reads that span the nick site

^f Alternative nucleotides in the consensus sequence are represented by IUPAC ambiguity code symbols, where V= A or C or G, N = any nucleotide residue, D= A or G or T, R = A or G, sequences of nick regions represent the sequences in 3'-5' strand where the nicks are located

Figure S1. Confirmation of PR1 virion DNA end sequences. Chromatograms represent the results of Sanger sequencing of PR1 virion LTR ends with the use of primer complementary to the lower strand of L-LTR (OMLO811, right panel) or to the upper strand of R-LTR (OMLO810, left panel). The crossed picks represent so called „glorious A” - a single false A nucleotide that is added at the physical end of DNA fragment by the sequencing DNA polymerase and does not represent any real sequence. Capital letters below the chromatograms represent redundant sequences at ends of LTRs, while small letters represent non-redundant genome parts at one border of each LTR. Regions matching to primers are shown schematically above the chromatograms at the scheme of PR1 virion DNA.

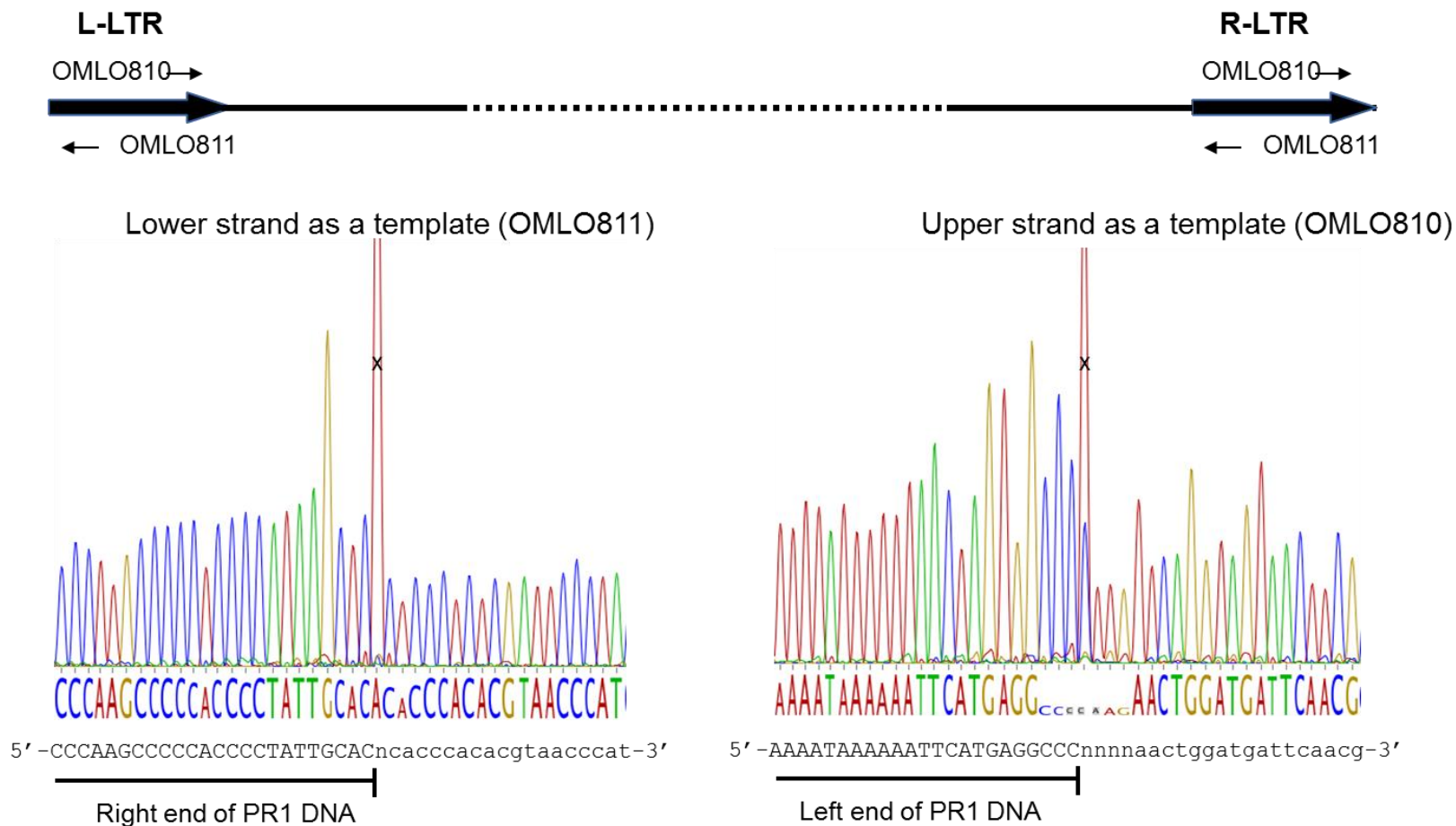


Figure S2. Analysis of the selected nick regions in PR1 virion DNA. Chromatograms represent the results of Sanger sequencing of nick region 1 and 9 (left panel) and 6 (right panel) with the use of primers complementary to the upper (OMLO812 and OMLO814, respectively) or to the lower strand (OMLO813 and OMLO815, respectively) of the virion DNA. The crossed picks represent the so called „glorious A” - a single false A nucleotide that is added at the physical end of DNA fragment by the sequencing DNA polymerase and does not represent any real sequence. The numbering of nicks is as in Table S4. Sequences of primers are listed in the Materials and Methods section.

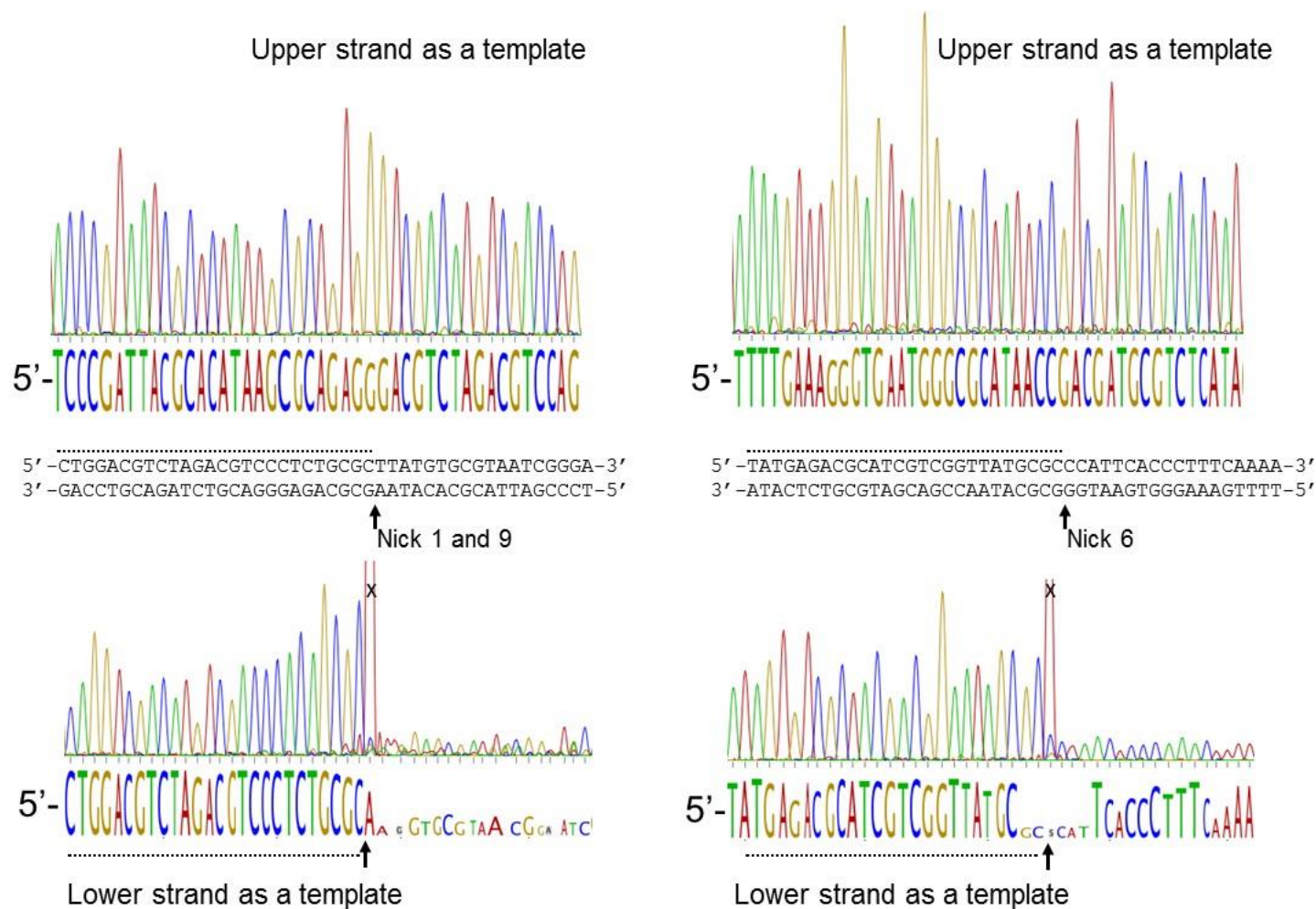


Figure S3. Orthologous clusters in *T5virus*. Comparisons of orthologous clusters of *T5virus* with *Providencia* virus PR1 (encoding 157 CDSs). The following *T5virus* were included: *Escherichia* virus T5 (encoding 157 CDSs), *Escherichia* virus AKFV33 (encoding 162 CDSs), *Escherichia* virus DT57C (encoding 132 CDSs), *Salmonella* virus Stitch (encoding 165 CDSs) and *Salmonella* virus Shivani (encoding 154 CDSs). Gene products at the terminal repeats were only considered once, for easy of interpretation of intra and interspecies protein relationships. The orthologous clusters were conducted using OrthoVenn server which only allows a maximum number 6 species to be analysed.

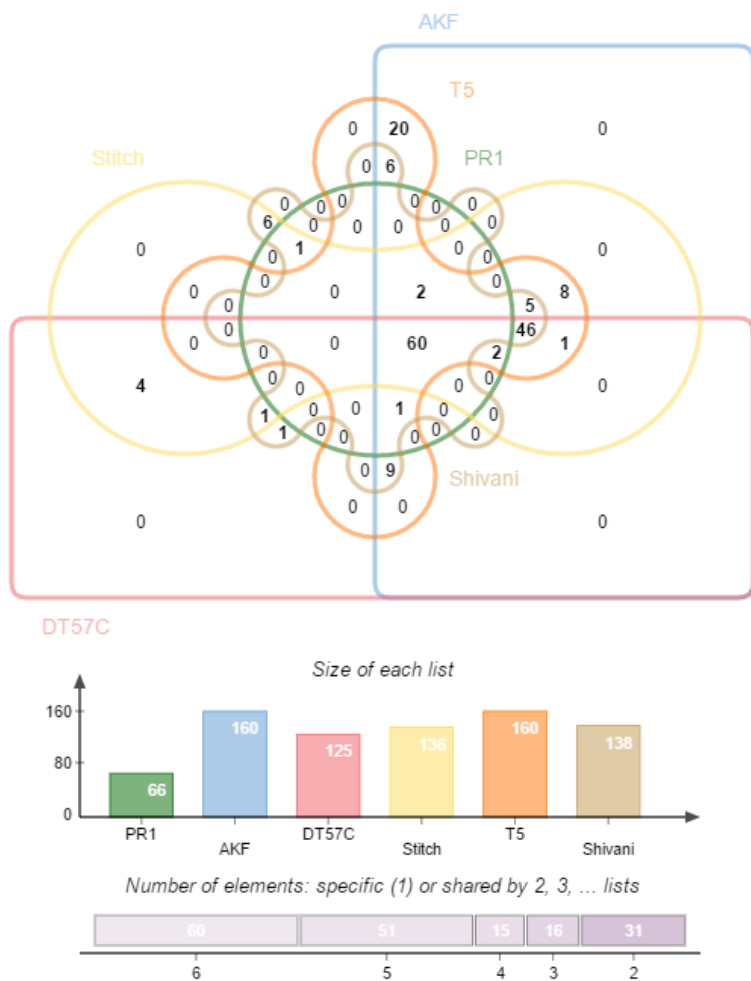
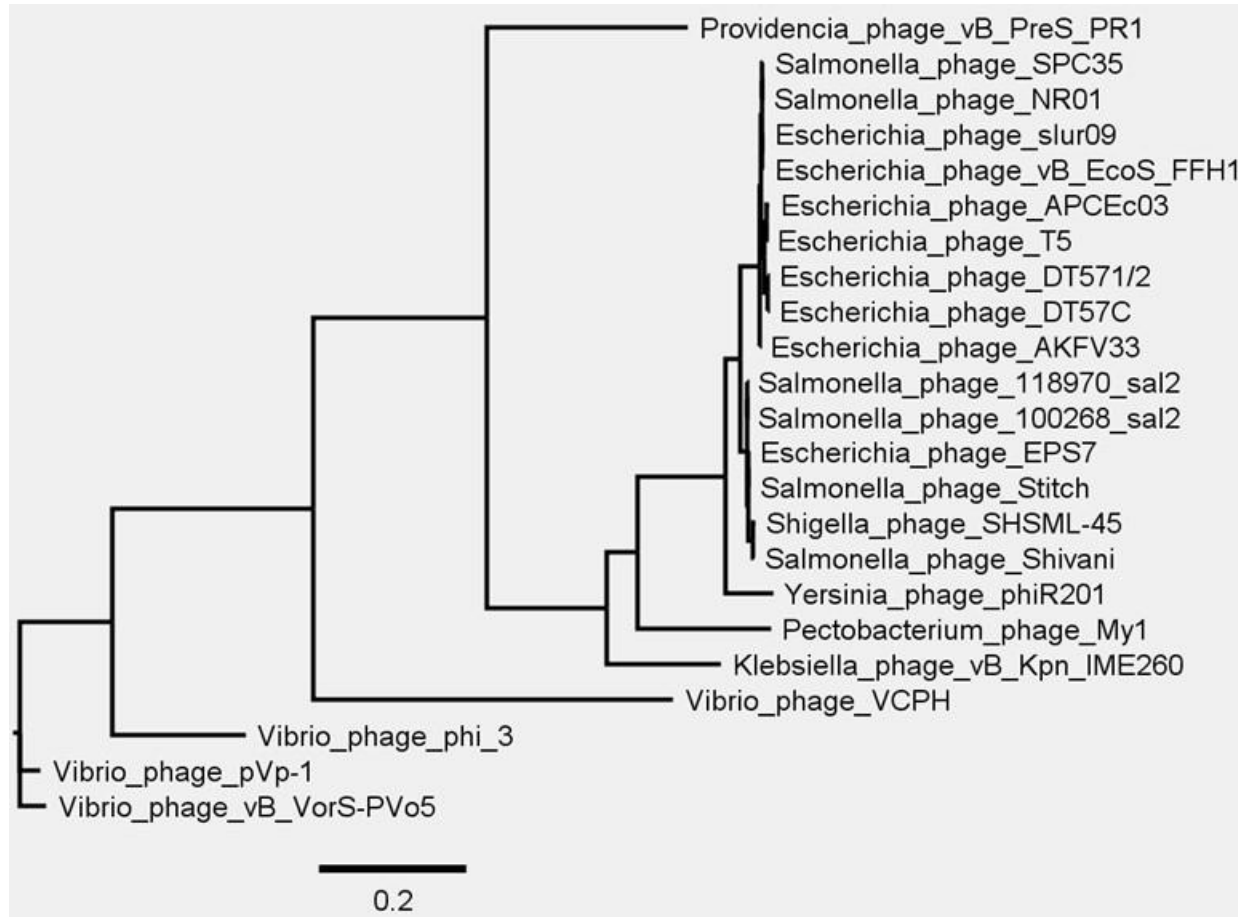
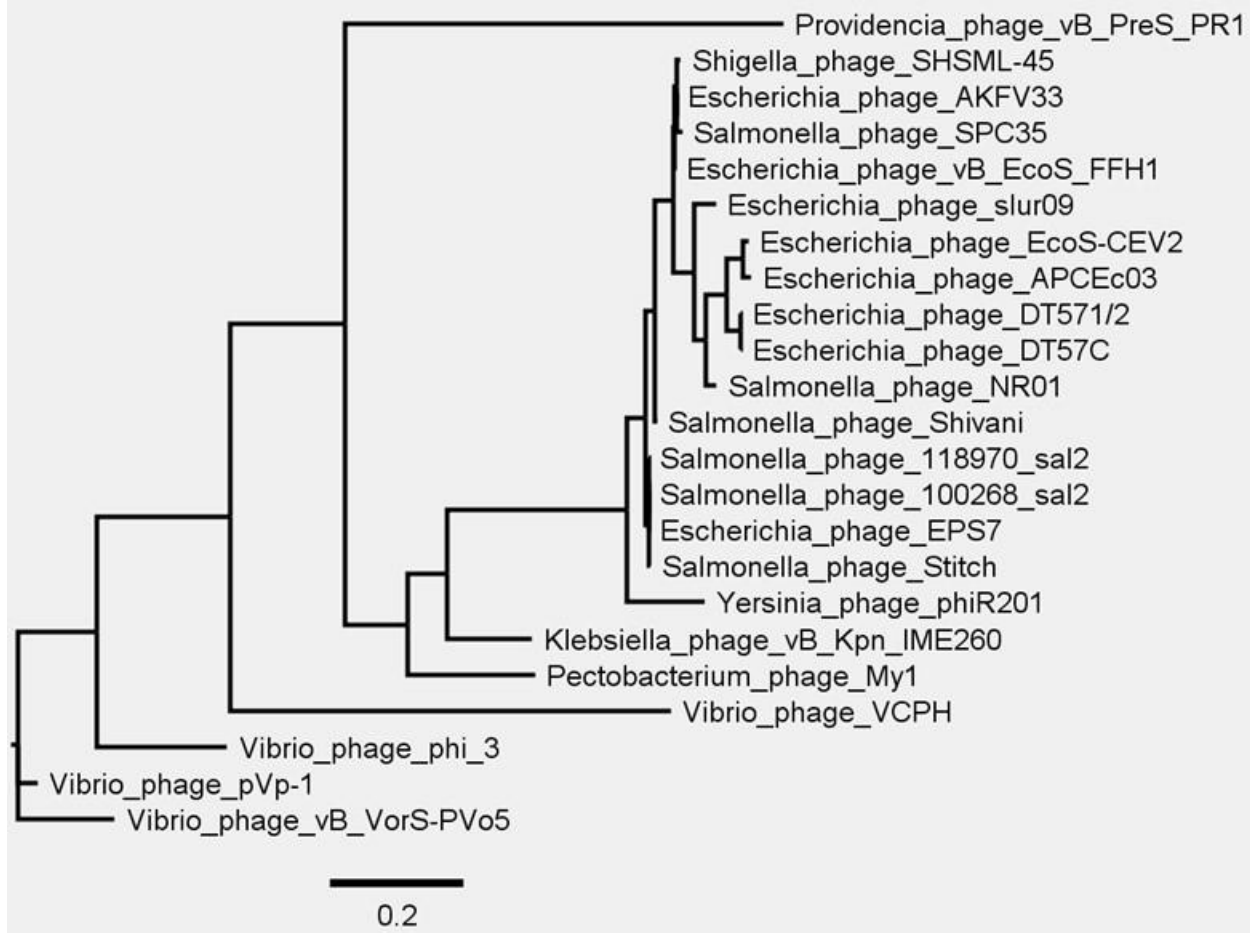


Figure S4. Phylogenetic analysis of phage proteins. DNA polymerase, major capsid protein, tail measure protein and large terminase subunit of PR1 homologs identified using BLASTp in GenBank and phages were used to generate phylogenetic trees using the "one click" mode at phylogeny.fr. The trees were exported in Newick format and tree produced using FigTree.

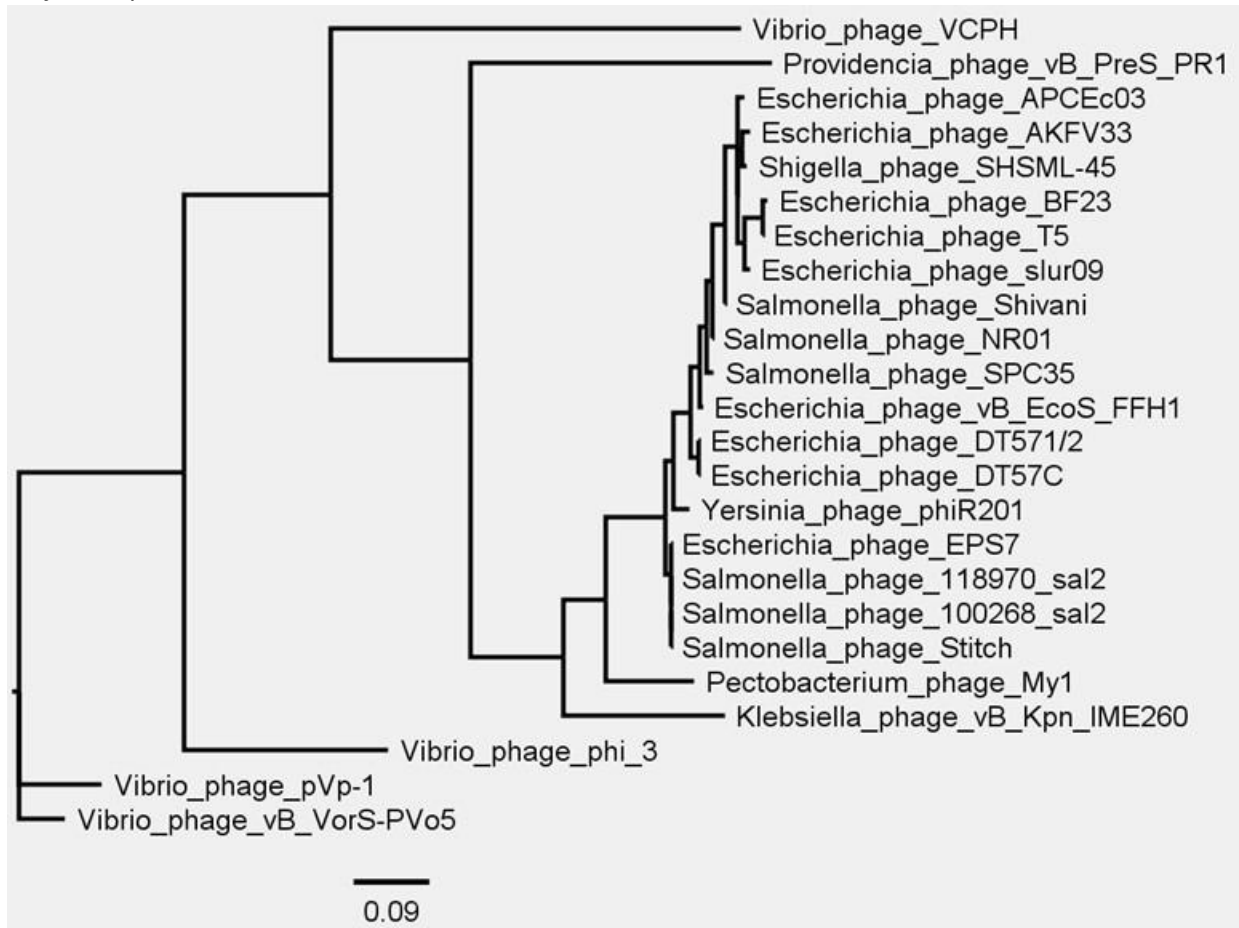
A. DNA polymerase



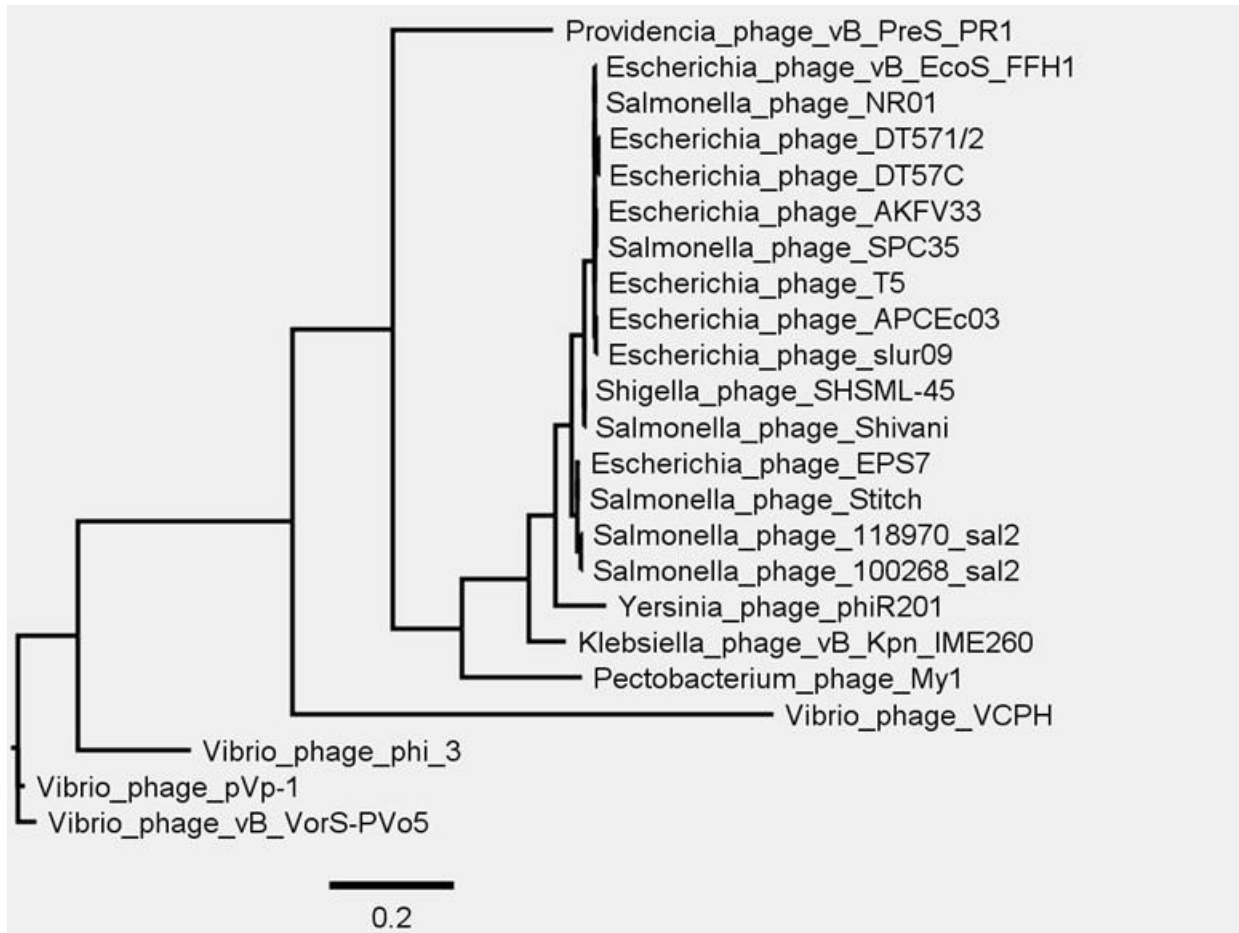
B. Major capsid protein



C. Major tail protein



D. Terminase, large subunit



References

1. Davison J. 2015. Pre-early functions of bacteriophage T5 and its relatives. *Bacteriophage* 5:e1086500.
2. Zivanovic Y, Confalonieri F, Ponchon L, Lurz R, Chami M, Flayhan A, Renouard M, Huet A, Decottignies P, Davidson AR, Breyton C, Boulanger P. 2014. Insights into Bacteriophage T5 Structure from Analysis of Its Morphogenesis Genes and Protein Components. *J Virol* 88:1162-1174.
3. Wang J, Jiang Y, Vincent M, Sun Y, Yu H, Bao Q, Kong H, Hu S. 2005. Complete genome sequence of bacteriophage T5. *Virology* 332:45-65.