

Supplementary Materials

Screening of bacteriophage encoded toxic proteins with a next generation sequencing based assay

Jutta Kasurinen¹, Cindy M. Spruit^{1,†}, Anu Wicklund^{1,2}, Maria I. Pajunen^{1,‡}, and Mikael Skurnik^{1,2,*,§}

¹ Department of Bacteriology and Immunology, Medicum, Human Microbiome Research Program, Faculty of Medicine, University of Helsinki, 00290 Helsinki, Finland; jutta.kasurinen@helsinki.fi (J.K.), c.m.spruit@uu.nl (C.M.S.); anumaria.wicklund@gmail.com (A.W.); maria.pajunen@helsinki.fi (M.I.P.); mikael.skurnik@helsinki.fi (M.S)

² Division of Clinical Microbiology, HUSLAB, University of Helsinki and Helsinki University Hospital, 00290 Helsinki, Finland

[†] Present address: Department of Chemical Biology & Drug Discovery, Utrecht Institute for Pharmaceutical Sciences, Utrecht University, 3584 CG Utrecht, the Netherlands

[‡] Shared senior authorship

^{*} Correspondence: mikael.skurnik@helsinki.fi; Tel.: +358 50 3360981

Supplementary Figures

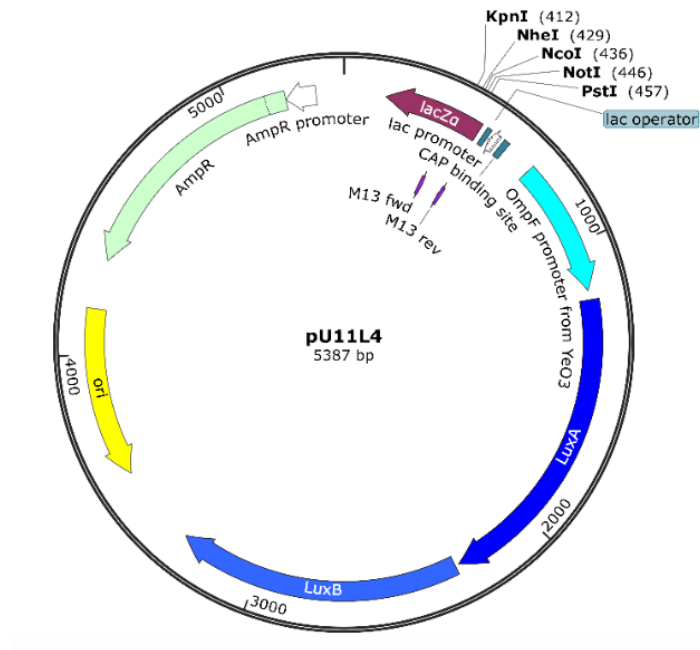


Figure S1. Map of plasmid pU11L4 [1].

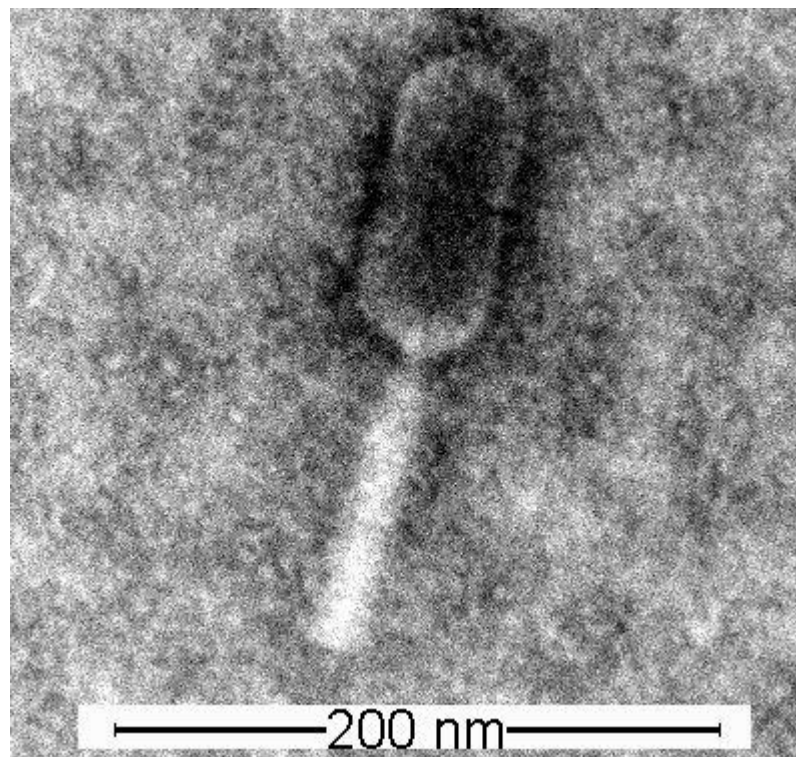


Figure S2. Transmission electron micrograph of a negatively stained fHy-Eco03 particle (capsid length 86 nm, tail length 103 nm, phage length 208 nm).



Figure S3. Plaque morphology of *Escherichia coli* strain #5509 infected with fHy-Eco03.

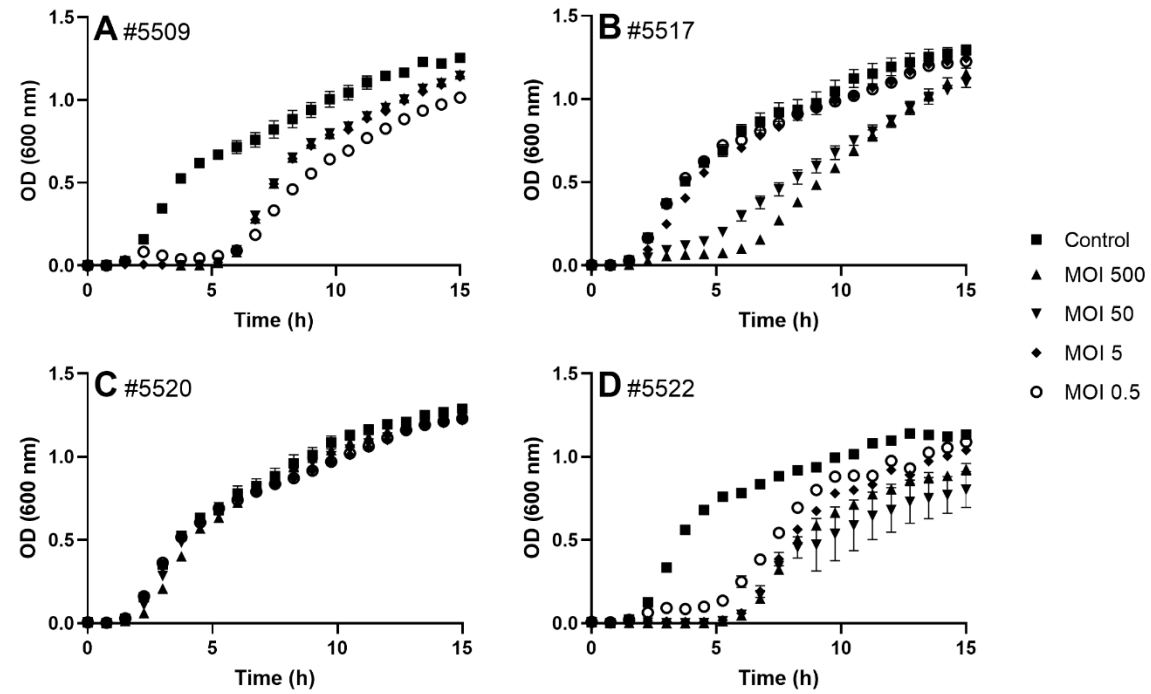


Figure S4. Growth curves of *Escherichia coli* strains #5509 (A), #5517 (B), #5520 (C), and #5522 (D) infected with fHy-Eco03. Bacteria were cultured in liquid LB medium at 37°C with different initial multiplicities of infections (MOIs). Each data point represents the average OD₆₀₀ for three to four replicates and error bars represent the standard deviation (SD).

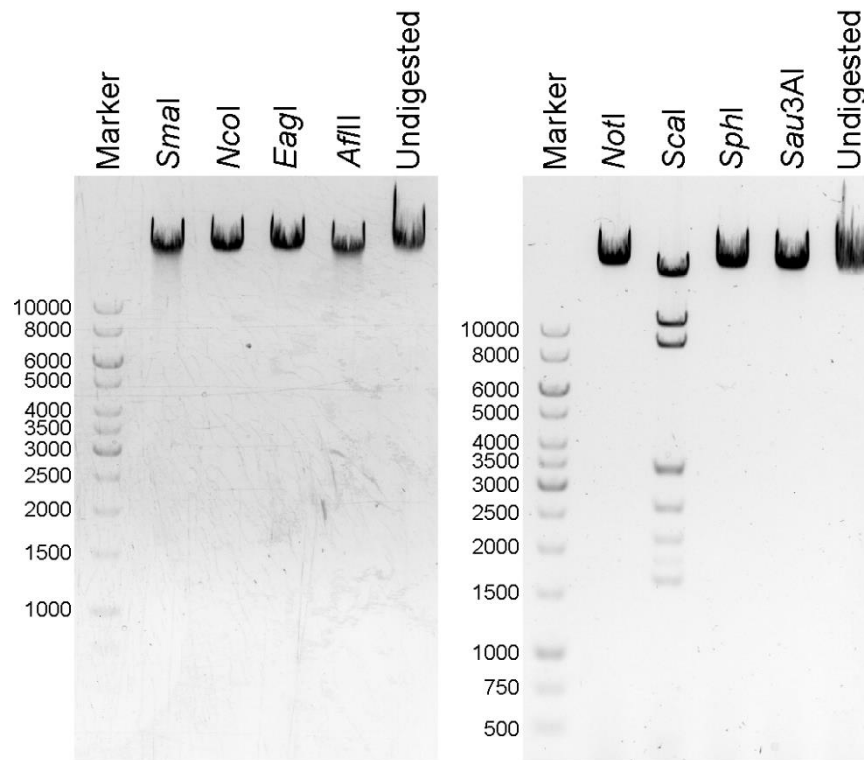
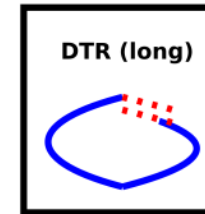
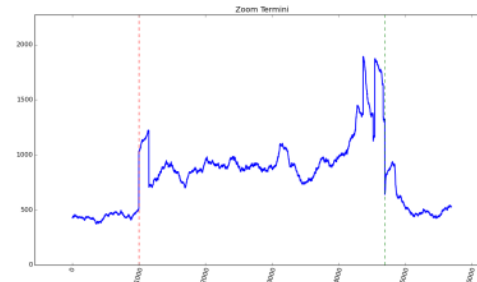


Figure S5. Restriction enzyme digestion analysis of fHy-Eco03 DNA. The DNA was digested with restriction endonucleases *AflIII*, *EagI*, *NcoI*, *NotI*, *Sau3AI*, *ScaI*, *SmaI*, and *SphI* and the digested samples were loaded to the gel as indicated, along with undigested phage DNA and 1 kb DNA ladder (Marker). The *in silico* ScaI-digested genomic sequence gives the following fragments (bp): 24,830; 10,405; 8,539; 3,337; 3,260 (left end); 2,531; 2,074; 1,596; 439 (right end). For the used restriction enzymes, the numbers of predicted recognition sites in the genome are as follows:

Enzyme	Recognition sequence	<i>n</i> sites	Methylation sensitivity / dependence (https://enzymefinder.neb.com/)
<i>AflIII</i>	CTTAAG	10	None/None
<i>EagI</i>	CGGCCG	5	CpG blocked/None
<i>NcoI</i>	CCATGG	7	None/None
<i>NotI</i>	GCGGCCGC	1	CpG variable/None
<i>Sau3AI</i>	GATC	9	Some block/None
<i>ScaI</i>	AGTACT	8	EcoBI: blocked/None
<i>SmaI</i>	CCCGGG	3	CpG blocked/None
<i>SphI</i>	GCATGC	14	EcoBI: blocked/None

fHyEco PhageTerm Analysis



PhageTerm Method

Ends	Left (red)	Right (green)	Permuted	Orientation	Class	Type
Redundant	47595	51293	No	NA	DTR (long)	T5

*Direct Terminal Repeats: 3699 bp

Strand	Location	T	pvalue	T (Start. Pos. Cov. / Whole Cov.)
+	47595	0.66	1.47e-114	
	51118	0.40	2.11e-26	
	41280	0.10	1.30e-04	
	23058	0.06	1.00e+00	
	33934	0.06	2.11e-01	
-	51293	0.68	1.03e-195	
	51117	0.59	3.89e-149	
	51453	0.07	1.67e-07	
	35956	0.06	9.95e-01	
	23478	0.06	1.00e+00	

Figure S6. Detection of genome ends of fHy-Eco03. PhageTerm output (according to the PhageTerm method) showing major peaks in the ends of the genome on both positive and negative strands.



Figure S7. Genomic map of fHy-Eco03. Geneious Prime (version 2021.0.3) program was used for visualization. The genes encoding structural or phage particle associated proteins (PPAPs) are drawn in brown, nucleic acid metabolism associated enzymes in red, cell lysis associated proteins in yellow, endonuclease in green, and hypothetical proteins in grey. PPAP, phage particle associated protein; HP, hypothetical protein.

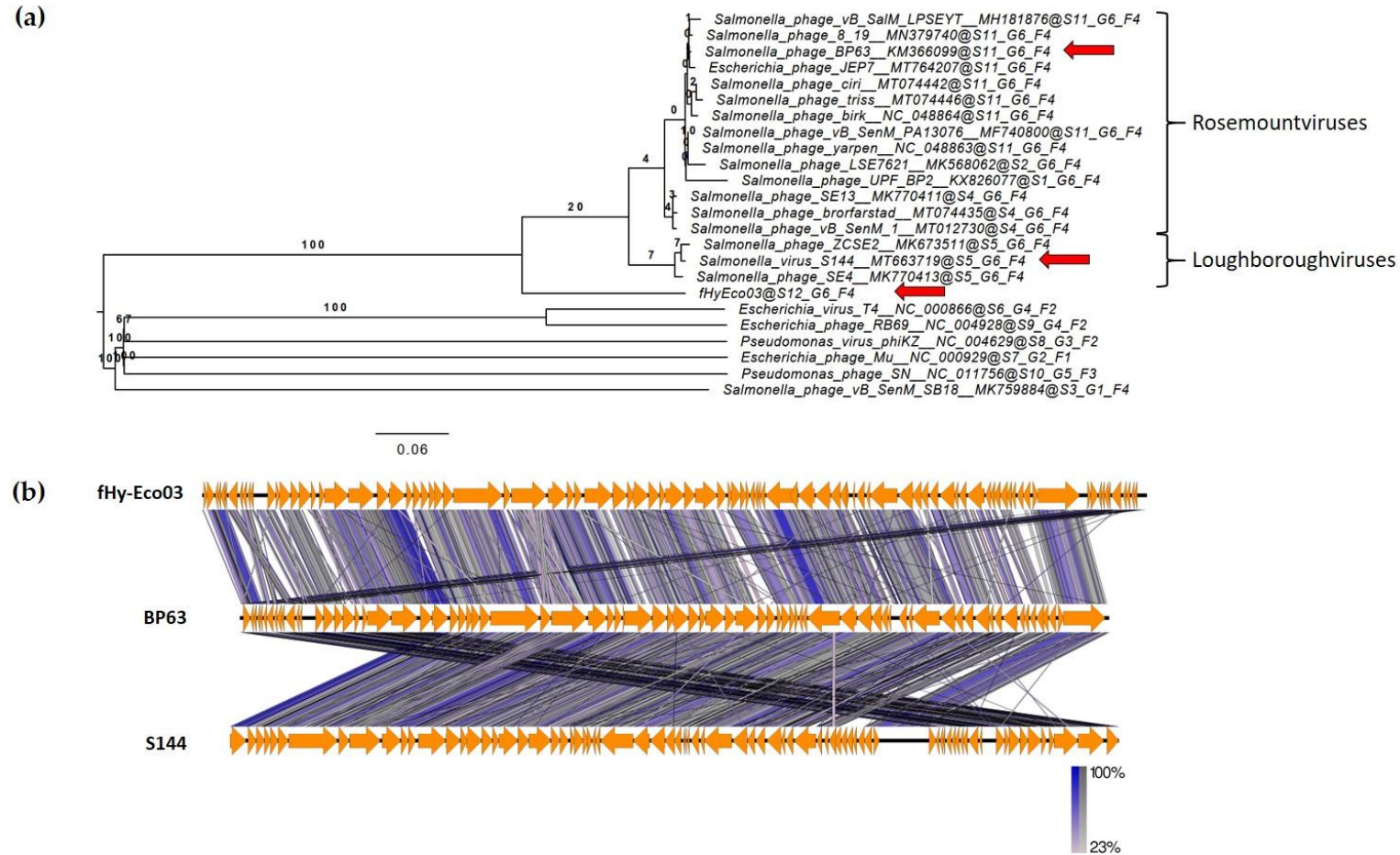


Figure S8. Genomic comparison of fHy-Eco03 (a) The phylogenetic genome-BLAST distance phylogeny (GBDP-D0) using VICTOR tool. Salmonella phage BP63 is the type phage for Rosemountviruses, and Salmonella phage ZCSE2 is the type phage for Loughboroughviruses, present in the ICTV virus taxonomy (release 2019). Some additional Myoviruses, especially those identified in the HHPRED analysis (Table S2) were included in the analysis. (b) An EasyFig generated comparison of the genomes of phages fHy-Eco03 (top), BP63 (middle) and S144 (bottom), indicated in panel (a) with red arrows, with different shades of blue indicating the identities between the genomes.

Supplementary Tables

Table S1. *Escherichia coli* strains used to determine the host range of fHy-Eco03. All the *E. coli* strains were isolated from human patient samples at the Clinical Microbiology Laboratory of the hospital district of Helsinki and Uusimaa, Finland. The Skurnik laboratory strain collection storage numbers are shown. Strains that produce extended-spectrum beta-lactamase (ESBL) are indicated. Strains that were infected by phage fHy-Eco03 are indicated as follows: ++, clear strong lysis; +/-, lysis but low EOP; -, no infection. The sensitivity of the strains indicated **in bold** to fHy-Eco03 was confirmed by determining infection growth curves using the Bioscreen C incubator (see Figure S4).

Storage #	Strain	Origin	Antibiotic resistance	Sensitivity to fHy-Eco03
5506	123794	Blood		–
5507	123789	Blood		–
5509	123781	Blood		++
5510	123780	Blood		–
5512	123770	Blood		–
5517	123745	Blood		++
5519	123741	Blood		–
5520	123740	Blood		+/- Turbid plaques
5521	123738	Blood		–
5522	123734	Blood		Turbid plaques
5625	13ME7768	Appendix abscess		–
	ME07799	Gall bladder fluid		–
5641	UB14789	Urine		–
5768	13MD3432	Rectal secretion	ESBL	–
	UB15006	Urine	ESBL	–
	KK08907	Amniotic fluid		–
	KK08908	Sacrum wound		–
	KP09927	Surgical wound		–
	UC14791	Urine	ESBL	–

5626	UF0545	Urine	ESBL	–
5759	EL3622	Feces	ESBL	Turbid plaques
5760	EL3615	Feces	ESBL	Turbid plaques
5761	W6215	Blood	ESBL	Turbid plaques
5762	EL3628	Feces	ESBL	–
5763	UC14886	Urine	ESBL	–
5764	UD14891	Urine	ESBL	–
	KP09910	Abscess		–
5627	UC 14827	Urine	ESBL	–
5628	UE14765	Secretion	ESBL	–
5765	UF05451	Urine	ESBL	–
5629	ME8280	Abdominal puncture		–
5766	13KP10517	Abscess in gluteal cleft		–
5767	13KP10501	Foot surgical wound		–
5630	TS07654	Thigh wound		–
5631	KK4340	Foot sole wound		–
5632	UB15026	Rectum mucus	ESBL	–
	UB15626	Urine, catheter	ESBL	–
	TS7670	Surgical wound		–
5633	EL3851	Rectum mucus	ESBL	–
5634	MD3414	Rectum mucus	ESBL	–
5635	UA15942	Urine	ESBL	–
	UA16057	Urine	ESBL	Turbid plaques
	UB15620	Urine	ESBL	–
5636	UB15822	Urine	ESBL	Turbid plaques

	UD15570	Urine	ESBL	–
	UE15406	Urine	ESBL	–
	EL3867	Rectum mucus	ESBL	Turbid plaques
5637	EL3870	Feces	ESBL	Turbid plaques
5638	EL3873	Rectum mucus	ESBL	–
5639	EL3875	Urine	ESBL	–

Table S2. Annotation of the predicted gene products of fHy-Eco03. The genes, gene location, gene size, and protein length are listed. LC-MS/MS analysis was performed to identify phage particle associated proteins (PPAPs). The best HHpred hits (with a probability above 50% and E-value below 1) are included [2]. A BLASTP search against the non-redundant protein sequences database (release update from February 6, 2021) was performed for every predicted gene product and the two results with the lowest E-values were recorded including query coverage, amino acid identity, and the accession numbers of the closest homologs. Note that due to the 3229 bp terminal repeats genes *g01* – *g11R* have two genomic locations. The fHy-Eco03 genome sequence with annotation has been deposited in the NCBI nucleotide database (GenBank) under accession number MW602648.

fHy-Eco03 genome features						The best HHpred hits (Probability percentage/E-value)	BLASTP analysis				
Gene	Start position	End position	Gene size (bp)	Protein size (aa)	PPAP		The best hits	Query coverage %	E-value	Amino acid identity %	Accession number
<i>g01</i>	114	311	198	65	no	No significant hits	hypothetical protein zoltan_21 [<i>Salmonella</i> phage zoltan]	93	0.007	40.32	QIN98827.1
	53426	53623					hypothetical protein ciri_21 [<i>Salmonella</i> phage ciri]	93	0.007	40.32	QIN98756.1
<i>g02</i>	311	736	426	141	yes	1OB8_B, Holliday-junction resolvase { <i>Saccharolobus solfataricus</i> } (99.6%/1.3*10 ⁻¹⁴)	putative resolvase [<i>Salmonella</i> virus S144]	100	2e-77	74.26	QMV47894.1
	53623	54048					hypothetical protein HOV36_gp57 [<i>Salmonella</i> phage ZCSE2]	100	1e-76	73.27	YP_009821772.1
<i>g03R</i>	962	777	186	61	no	No significant hits	hypothetical protein triss_20 [<i>Salmonella</i> phage triss]	100	8e-34	88.52	QIN99040.1
	54274	54089					hypothetical protein HWB04_gp70 [<i>Salmonella</i> phage UPF_BP2]	100	2e-33	88.52	YP_009831725.1
<i>g04R</i>	1138	959	180	59	no	No significant hits	hypothetical protein BJD50_gp03 [<i>Salmonella</i> phage BP63]	100	7e-36	98.31	YP_009302922.1
	54450	54271					hypothetical protein JEP7_56 [<i>Escherichia</i> phage JEP7]	100	2e-35	98.31	QOC55413.1
<i>g05R</i>	1383	1138	246	81	no	No significant hits	hypothetical protein HOV36_gp60 [<i>Salmonella</i> phage ZCSE2]	100	2e-39	77.78	YP_009821775.1
	54695	54450					hypothetical protein HWC20_gp76 [<i>Salmonella</i> phage SE4]	92	4e-39	82.67	YP_009845656.1
<i>g06R</i>	1492	1373	120	39	no	No significant hits	hypothetical protein [<i>Salmonella</i> phage 8-19]	94	1e-08	67.57	QFR58182.1
	54804	54685					hypothetical protein HWC19_gp07 [<i>Salmonella</i> phage SE13]	97	2e-08	65.79	YP_009845514.1
<i>g07R</i>	2122	1550	573	190	yes	No significant hits	hypothetical protein HOV36_gp67 [<i>Salmonella</i> phage ZCSE2]	96	1e-74	63.59	YP_009821782.1
	55434	54862					hypothetical protein [<i>Salmonella</i> virus S144]	96	3e-74	63.04	QMV47903.1

g08R	2453	2253	201	66	no	No significant hits	hypothetical protein HOV36_gp63 [<i>Salmonella</i> phage ZCSE2]	98	4e-09	41.54	YP_009821778.1
	55765	55565					hypothetical protein [<i>Salmonella</i> virus S144]	98	5e-09	41.54	QMV47899.1
g09R	2684	2457	228	75	no	No significant hits	hypothetical protein brorfarstad_14 [<i>Salmonella</i> phage brorfarstad]	100	4e-30	66.67	QIN98305.1
	55996	55769					hypothetical protein vBSenM1_37 [<i>Salmonella</i> phage vB_SenM-1]	100	7e-30	66.67	QJQ40037.1
g10R	2930	2754	177	58	no	No significant hits	hypothetical protein [<i>Salmonella</i> virus S144]	100	1e-22	79.31	QMV47901.1
	56242	56066					hypothetical protein HWC20_gp71 [<i>Salmonella</i> phage SE4]	98	1e-22	82.46	YP_009845651.1
g11R	3918	4262	171	56	no	No significant hits	hypothetical protein HWD20_gp09 [<i>Salmonella</i> phage birk]	83	7e-04	38.30	YP_009857669.1
	56412	56242					hypothetical protein vBSalMLPSEYT_00001 [<i>Salmonella</i> phage vB_SalM-LPSEYT]	83	0.002	38.30	AWW14697.1
g12	4312	4461	495	164	yes	No significant hits	hypothetical protein nenneke_8 [<i>Salmonella</i> phage nenneke]	98	2e-23	40.00	QIN98440.1
							hypothetical protein rivia_8 [<i>Salmonella</i> phage rivia]	98	8e-23	38.79	EBY8042318.1
g13	4461	4697	237	78	no	No significant hits	No significant similarity found				
g14	4697	5368	672	223	yes	4PSS_A, Dihydrofolate reductase [<i>Escherichia coli</i>] (99.91%/3.9*10 ⁻²³)	dihydrofolate reductase [<i>Salmonella</i> phage vB_SalM-LPSEYT]	100	3e-92	59.82	AWW14768.1
							dihydrofolate reductase [<i>Salmonella</i> phage SE13]	100	1e-91	59.82	YP_009845524.1
g15	5365	5871	507	168	no	4MQB_B, Thymidylate kinase [<i>Staphylococcus aureus</i>] (99.71%/5.5*10 ⁻¹⁵)	thymidylate kinase [<i>Salmonella</i> phage 8-19]	100	1e-98	80.95	QFR58193.1
							thymidylate kinase [<i>Salmonella</i> phage birk]	100	3e-98	80.95	YP_009857666.1
g16	5868	6590	723	240	yes	No significant hits	hypothetical protein HOV36_gp73 [<i>Salmonella</i> phage ZCSE2]	95	4e-15	30.80	YP_009821788.1
g17	6590	6880	291	96	no	No significant hits	hypothetical protein [<i>Salmonella</i> virus S144]	100	5e-37	59.38	QMV47832.1
							hypothetical protein HOV36_gp74 [<i>Salmonella</i> phage ZCSE2]	100	2e-36	58.33	YP_009821789.1
g18	7076	7399	324	107	no	No significant hits	hypothetical protein HWD20_gp03 [<i>Salmonella</i> phage birk]	87	3e-41	72.34	YP_009857663.1
							hypothetical protein HWC19_gp21 [<i>Salmonella</i> phage SE13]	87	3e-41	71.28	YP_009845528.1
	7389	8840	1452	483	yes		terminase large subunit [<i>Salmonella</i> phage SE4]	100	0.0	83.02	YP_009845640.1

g19						3CPE_A, Large terminase, gp17 { <i>Escherichia</i> virus T4} (100%/3.5*10 ⁻³²)	hypothetical protein HOV36_gp76 [<i>Salmonella</i> phage ZCSE2]	100	0.0	82.82	YP_009821791.1
g20	8842	10389	1548	515	yes	2JES_K, Portal protein { <i>Bacillus</i> phage SSP1} (99.96%/1.1*10 ⁻²⁶)	portal protein [<i>Salmonella</i> phage ciri]	99	0.0	77.80	QIN98736.1
							portal protein [<i>Salmonella</i> phage BP63]	99	0.0	77.61	YP_009302940.1
g21	10566	11282	717	238	yes	21C6_A, Sin Nombre Virus Nucleocapsid Protein (94.9%/0.93)	hypothetical protein HOV36_gp78 [<i>Salmonella</i> phage ZCSE2]	100	1e-109	81.09	YP_009821793.1
							scaffold protein [<i>Salmonella</i> phage SE4]	99	2e-131	81.36	YP_009845638.1
g22	11285	12253	969	322	yes	6R3A_G, Major capsid protein { <i>Bacillus</i> phage SSP1} (100%/8.3*10 ⁻³⁷)	major capsid protein [<i>Salmonella</i> phage brunost]	100	0.0	87.38	QIN98734.1
							major capsid protein [<i>Salmonella</i> phage BP63]	83	0.0	87.38	YP_009302942.1
g23	12326	12694	369	122	no	5H0M_A, HNH endonuclease { <i>Geobacillus</i> virus E2} (99.17%/3*10 ⁻¹¹)	HNH endonuclease [<i>Salmonella</i> phage ZCSE2]	95	8e-50	68.10	YP_009821717.1
							hypothetical protein [<i>Klebsiella</i> phage NL_ZS_3]	95	1e-32	55.66	QNN97356.1
g24	12730	13224	495	164	yes	5A21_D, Head-to-tail interface { <i>Bacillus</i> phage SSP1} (96.05%/0.27)	hypothetical protein HWC20_gp56 [<i>Salmonella</i> phage SE4]	100	2e-79	69.28	YP_009845636.1
							putative head-to-tail connector complex 1 [<i>Salmonella</i> virus S144]	99	4e-78	68.07	QMV47838.1
g25	13226	13693	468	155	yes	No significant hits	putative head-to-tail connector complex 2 [<i>Salmonella</i> virus S144]	99	2e-52	52.56	QMV47839.1
							hypothetical protein HOV36_gp04 [<i>Salmonella</i> phage ZCSE2]	99	3e-52	52.56	YP_009821719.1
g26	13695	14096	402	133	no	5A21_F, Head completion protein, gp16 { <i>Bacillus</i> phage SSP1} (97.96%/4.0*10 ⁻⁴)	hypothetical protein HWC20_gp54 [<i>Salmonella</i> phage SE4]	98	4e-52	64.89	YP_009845634.1
							hypothetical protein HWC19_gp28 [<i>Salmonella</i> phage SE13]	98	2e-51	64.12	YP_009845535.1
g27	14020	14553	534	177	yes	5A21_G, Tail-to-head joining protein, gp17 { <i>Bacillus</i> phage SSP1} (96.59%/0.075)	hypothetical protein BJD50_gp27 [<i>Salmonella</i> phage BP63]	92	8e-60	57.67	YP_009302946.1
							hypothetical protein HWB04_gp48 [<i>Salmonella</i> phage UPF_BP2]	92	2e-59	57.06	YP_009831703.1
g28	14556	15176	621	206	yes	No significant hits	hypothetical protein HWD19_gp65 [<i>Salmonella</i> phage yarpen]	97	9e-67	54.23	YP_009857656.1
							putative tail protein 1 [<i>Salmonella</i> virus S144]	99	3e-64	54.63	QMV47842.1
g29	15169	18132	2964	987	yes	No significant hits	putative tail protein 2 [<i>Salmonella</i> virus S144]	98	0.0	78.32	QMV47843.1
							tail protein I [<i>Salmonella</i> phage ZCSE2]	98	0.0	78.22	YP_009821723.1

g30	18199	18630	432	143	yes	5NGJ_B, Tail tube protein { <i>Escherichia</i> phage T5} (95.2%/0.78)	hypothetical protein [<i>Pseudobutyrvibrio</i> sp.]	88	4e-16	44.09	MBE5909634.1
							Ig-like domain-containing protein [<i>Cloacibacillus porcorum</i>]	61	7e-16	55.06	WP_168950698.1
g31	18670	20787	2118	705	yes	4SHS_A, Hoc Head outer capsid protein { <i>Escherichia</i> phage RB69} (96.42%/0.98)	hypothetical protein [<i>Salmonella</i> phage 8-19]	100	0.0	66.71	QFR58208.1
							putative Ig-like domain-containing protein [<i>Salmonella</i> phage UPF_BP2]	100	0.0	66.85	YP_009831699.1
g32	20874	22007	1134	377	yes	3CPE_A, Tail sheath protein, gp18 { <i>Escherichia</i> virus T4} (98.6%/2.4*10 ⁻⁵)	hypothetical protein HWC19_gp34 [<i>Salmonella</i> phage SE13]	100	0.0	75.20	YP_009845541.1
							hypothetical protein brorfarstad_62 [<i>Salmonella</i> phage brorfarstad]	100	0.0	74.93	QIN98353.1
g33	22019	22450	432	143	yes	6GKX_ R-type bacteriocin tube protein CD1364 { <i>Clostridium difficile</i> } (97.56%/2.9*10 ⁻³)	hypothetical protein HOV36_gp12 [<i>Salmonella</i> phage ZCSE2]	99	1e-90	88.73	YP_009821727.1
							DUF3277 domain-containing protein [<i>Escherichia coli</i>]	99	2e-89	88.03	WP_166428470.1
g34	22469	22927	459	152	yes	No significant hits	hypothetical protein HOV36_gp13 [<i>Salmonella</i> phage ZCSE2]	97	2e-53	58.78	YP_009821728.1
							hypothetical protein HWC20_gp46 [<i>Salmonella</i> phage SE4]	97	2e-53	59.46	YP_009845626.1
g35	23078	24769	1692	563	yes	6V8I_AF, Tape measure protein, gp57 { <i>Staphylococcus</i> virus 80alpha} (97.7%/0.2)	hypothetical protein brorfarstad_59 [<i>Salmonella</i> phage brorfarstad]	100	0.0	60.84	QIN98350.1
							tail tape measure protein [<i>Salmonella</i> phage vB_SenM-1]	100	0.0	61.27	QJQ40070.1
g36	24766	25665	900	299	yes	No significant hits	hypothetical protein vBSenM1_71 [<i>Salmonella</i> phage vB_SenM-1]	99	3e-152	72.82	QJQ40071.1
							hypothetical protein HWD20_gp61 [<i>Salmonella</i> phage birk]	99	2e-151	72.48	YP_009857721.1
g37	25662	26054	393	130	yes	No significant hits	hypothetical protein HWB04_gp39 [<i>Salmonella</i> phage UPF_BP2]	100	2e-59	63.85	YP_009831694.1
							hypothetical protein BJD50_gp38 [<i>Salmonella</i> phage BP63]	100	3e-59	63.08	YP_009302957.1
g38	25996	26943	948	315	yes	1WRU_A, Baseplate gp44 { <i>Enterobacteria</i> phage Mu} (99.9%/1.7*10 ⁻²¹)	hypothetical protein vBSalMLPSEYT_00050 [<i>Salmonella</i> phage vB_SalM-LPSEYT]	98	3e-165	72.35	AWW14746.1
							hypothetical protein BJD50_gp39 [<i>Salmonella</i> phage BP63]	98	5e-165	72.35	YP_009302958.1
g39	26945	27586	642	213	yes		hypothetical protein HWD19_gp54 [<i>Salmonella</i> phage yarpn]	98	2e-122	78.10	YP_009857645.1

						4RU3_A, Cell puncturing protein gp41 (<i>Pseudomonas</i> phage SN) (99.9%/1.1*10 ⁻²⁶)	baseplate protein [<i>Salmonella</i> phage vB_SalM-LPSEYT]	98	2e-122	78.10	AWW14745.1
g40	27595	27954	360	119	no	5IW9_B, Baseplate wedge protein, gp25 (<i>Escherichia</i> virus T4) (99.72%/1*10 ⁻¹⁵)	hypothetical protein HOV36_gp19 [<i>Salmonella</i> phage ZCSE2]	98	4e-65	99.25	YP_009821734.1
							hypothetical protein HWC20_gp40 [<i>Salmonella</i> phage SE4]	99	1e-64	98.49	YP_009845620.1
g41	27957	29114	1158	385	yes	5HX2_E, Baseplate wedge protein, gp6 (<i>Escherichia</i> virus T4) (100%/3.5*10 ⁻³⁰)	hypothetical protein brorfartstad_53 [<i>Salmonella</i> phage brorfartstad]	100	0.0	67.44	QIN98344.1
							hypothetical protein HWC19_gp43 [<i>Salmonella</i> phage SE13]	100	0.0	67.18	YP_009845550.1
g42	29107	29757	651	216	yes	No significant hits	hypothetical protein HOV36_gp21 [<i>Salmonella</i> phage ZCSE2]	100	7e-113	74.19	YP_009821736.1
							hypothetical protein vBSenM1_77 [<i>Salmonella</i> phage vB_SenM-1]	100	1e-110	73.27	QJQ40077.1
g43	29750	31099	1350	449	yes	5YVQ_A, Tail fiber protein S (<i>Enterobacteria</i> phage Mu) (96.59%/0.004)	putative tail fibers protein [<i>Salmonella</i> phage SE4]	100	0.0	67.26	YP_009845617.1
							tail fiber protein [<i>Salmonella</i> phage ZCSE2]	100	0.0	66.82	YP_009821737.1
g44	31096	31629	534	177	yes	5YVQ_B, Tail fiber assembly protein U (<i>Enterobacteria</i> phage Mu) (99.97%/2.2*10 ⁻²⁹)	hypothetical protein HOV36_gp23 [<i>Salmonella</i> phage ZCSE2]	98	3e-23	98.92	YP_009821738.1
							tail fibers protein [<i>Salmonella</i> phage SE4]	98	1e-20	96.24	YP_009845616.1
g45	31727	31972	246	81	yes	No significant hits	hypothetical protein HWC19_gp48 [<i>Salmonella</i> phage SE13]	96	9e-23	55.95	YP_009845555.1
							hypothetical protein vBSenM1_81 [<i>Salmonella</i> phage vB_SenM-1]	96	2e-22	55.95	QJQ40081.1
g46	31947	32510	564	187	yes	2ANV_A, Phage lysozyme (<i>Salmonella</i> phage P22) (99.81%/6.6*10 ⁻¹⁸)	putative endolysin [<i>Salmonella</i> virus S144]	97	5e-122	90.71	QMV47861.1
							lysozyme [<i>Salmonella</i> phage vB_SalM-LPSEYT]	97	1e-120	90.71	AWW14736.1
g47	32480	32830	351	116	yes	4LIN_G, Tail needle protein, gp26 (<i>Salmonella</i> phage P22) (95.71%/0.5)	putative IM-spanin [<i>Salmonella</i> virus S144]	100	6e-37	62.07	QMV47862.1
							hypothetical protein HOV36_gp27 [<i>Salmonella</i> phage ZCSE2]	100	7e-37	62.07	YP_009821742.1
g48	32814	33029	216	71	no	No significant hits	hypothetical protein rivia_44 [<i>Salmonella</i> phage rivia]	94	1e-29	74.63	QIN98547.1
							hypothetical protein HWD20_gp48 [<i>Salmonella</i> phage birk]	94	7e-28	73.13	YP_009857708.1

g49R	33320	333042	279	92	yes	No significant hits	hypothetical protein HWC19_gp52 [<i>Salmonella</i> phage SE13]	98	2e-54	89.01	YP_009845559.1
							ryanodine receptor [<i>Salmonella</i> phage vB_SalM-LPSEYT]	98	1e-52	85.71	AWW14733.1
g50R	33544	33320	225	74	no	4IJJ_B, C4-type zinc finger protein, DksA/TraR family (<i>Pseudomonas aeruginosa</i>)(99.7%/4.7*10 ⁻¹⁶)	hypothetical protein HWC19_gp53 [<i>Salmonella</i> phage SE13]	100	5e-42	89.19	YP_009845560.1
							DnaK suppressor protein [<i>Salmonella</i> phage vB_SalM-LPSEYT]	100	2e-40	85.14	AWW14732.1
g51R	33692	33537	156	51	no	No significant hits	No significant similarity found				
g52R	33964	33689	276	91	no	No significant hits	hypothetical protein HOV36_gp29 [<i>Salmonella</i> phage ZCSE2]	100	4e-49	99.51	YP_009821744.1
							hypothetical protein HWC19_gp54 [<i>Salmonella</i> phage SE13]	100	3e-41	99.35	YP_009845561.1
g53R	35953	33986	1968	655	yes	4X0Q_A, DNA polymerase theta (<i>Homo sapiens</i>)(100%/3.5*10 ⁻⁷⁰)	DNA polymerase [<i>Salmonella</i> phage vB_SenM_PA13076]	100	0.0	85.50	ATI16200.1
g53.1R	35916	35458	459	152	yes	No significant hits	DNA polymerase [<i>Salmonella</i> phage BP63]	100	0.0	85.50	YP_009302972.1
g54R	36947	35955	993	330	no	4TR6_B, DNA polymerase III subunit beta (<i>Bacillus subtilis</i>)(100%/8.9*10 ⁻⁴³)	DNA polymerase beta subunit [<i>Salmonella</i> phage LSE7621]	99	4e-166	66.87	QBJ00954.1
							DNA polymerase beta subunit [<i>Salmonella</i> phage vB_SenM_PA13076]	99	8e-166	66.57	ATI16199.1
g55R	37912	36983	930	309	yes	4EZ8_A, Thymidylate synthase (<i>Mus musculus</i>)(100%/1*10 ⁻⁵⁸)	thymidylate synthase [<i>Salmonella</i> phage ZCSE2]	100	5e-132	60.97	YP_009821747.1
							thymidylate synthase [<i>Salmonella</i> phage SE4]	100	2e-128	59.03	YP_009845606.1
g56R	38514	37912	603	200	yes	3NL9_A, NTP pyrophosphohydrolase (<i>Exiguobacterium sibiricum</i>)(99.96%/6.9*10 ⁻²⁹)	hypothetical protein HWC20_gp25 [<i>Salmonella</i> phage SE4]	99	4e-91	65.66	YP_009845605.1
							hypothetical protein HOV36_gp33 [<i>Salmonella</i> phage ZCSE2]	99	4e-90	64.14	YP_009821748.1
g57R	38899	38501	399	132	yes	No significant hits	hypothetical protein HWB04_gp21 [<i>Salmonella</i> phage UPF_BP2]	100	4e-25	42.54	YP_009831676.1
							hypothetical protein PA13076_52 [<i>Salmonella</i> phage vB_SenM_PA13076]	100	6e-25	42.54	ATI16196.1
g58	38888	39034	147	48	no	No significant hits	No significant similarity found				
g59R	39379	39254	126	41	no	No significant hits	No significant similarity found				
g60R	39964	39488	477	158	yes	No significant hits	hypothetical protein HWC19_gp60 [<i>Salmonella</i> phage SE13]	82	5e-68	91.60	YP_009845567.1

g61R	40290	40075	216	71	no	No significant hits	hypothetical protein HOV36_gp36 [<i>Salmonella</i> phage ZCSE2]	97	9e-60	81.82	YP_009821751.1
							hypothetical protein HOV36_gp37 [<i>Salmonella</i> phage ZCSE2]	100	2e-19	50.70	YP_009821752.1
							hypothetical protein HWC20_gp21 [<i>Salmonella</i> phage SE4]	100	4e-19	50.70	YP_009845601.1
g62R	41941	40268	1674	557	yes	2OCA_A, ATP-dependent helicase, uvsW [<i>Escherichia</i> virus T4] (100%/2.6*10 ⁻³³)	ATP-dependent RNA helicase [<i>Salmonella</i> phage ZCSE2]	99	0.0	84.41	YP_009821753.1
							putative ATP-dependent helicase [<i>Salmonella</i> virus S144]	99	0.0	84.23	QMV47876.1
g63R	42877	42005	873	290	yes	3SM4_C, Exonuclease [<i>Escherichia</i> virus Lambda] (100%/5.3*10 ⁻¹¹)	hypothetical protein HWC20_gp19 [<i>Salmonella</i> phage SE4]	98	9e-147	69.79	YP_009845599.1
							hypothetical protein HOV36_gp40 [<i>Salmonella</i> phage ZCSE2]	98	5e-144	68.40	YP_009821755.1
g64R	43388	42840	549	182	no	1U3E_M, HNH homing endonuclease [<i>Bacillus</i> virus SPO1] (93.91%/1.6*10 ⁻¹²)	HNH endonuclease [<i>Acidibacillus sulfuroxidans</i>]	92	2e-25	41.14	WP_109431645.1
							homing endonuclease [<i>Enterobacteria</i> phage JenK1]	92	4e-25	41.14	YP_009219324.1
g65R	43806	43357	450	149	no	2W4L_F, deoxycytidylate deaminase [<i>Homo sapiens</i>] (99.93%/4.4*10 ⁻²⁴)	hypothetical protein HOV36_gp41 [<i>Salmonella</i> phage ZCSE2]	91	6e-41	52.48	YP_009821756.1
							deoxycytidylate deaminase [<i>Salmonella</i> phage birk]	91	3e-40	49.29	YP_009857696.1
g66R	444393	43866	528	175	yes	No significant hits	hypothetical protein HWD20_gp35 [<i>Salmonella</i> phage birk]	100	2e-88	71.82	YP_009857695.1
							hypothetical protein brorfstad_33 [<i>Salmonella</i> phage brorfstad]	100	2e-87	69.61	QIN98324.1
g67R	45388	44495	894	297	yes	1C9K_C, Adenosyl-cobinamide kinase [<i>Salmonella typhimurium</i>] (99.33%/1.9*10 ⁻¹¹)	putative P-loop containing nucleoside triphosphatehydrolase [<i>Salmonella</i> virus S144]	93	8e-162	76.70	QMV47880.1
							hypothetical protein HWC20_gp16 [<i>Salmonella</i> phage SE4]	93	1e-161	77.14	YP_009845596.1
g68R	45619	45389	231	76	no	No significant hits	hypothetical protein HWC20_gp15 [<i>Salmonella</i> phage SE4]	100	8e-23	57.89	YP_009845595.1
							hypothetical protein [<i>Salmonella</i> virus S144]	100	1e-22	59.21	QMV47881.1
g69R	46099	45677	423	140	no	No significant hits	hypothetical protein HWC20_gp14 [<i>Salmonella</i> phage SE4]	95	2e-19	36.17	YP_009845594.1

							hypothetical protein HOV36_gp45 [<i>Salmonella</i> phage ZCSE2]	94	4e-19	36.69	YP_009821760.1
g70R	47200	46148	1053	350	yes	No significant hits	hypothetical protein HOV36_gp46 [<i>Salmonella</i> phage ZCSE2]	100	5e-97	47.21	YP_009821761.1
							hypothetical protein [<i>Salmonella</i> virus S144]	100	4e-96	46.93	QMV47883.1
g71R	47572	47273	300	99	no	No significant hits	hypothetical protein HWC20_gp12 [<i>Salmonella</i> phage SE4]	95	1e-35	61.05	YP_009845592.1
							hypothetical protein HOV36_gp47 [<i>Salmonella</i> phage ZCSE2]	96	1e-35	60.42	YP_009821762.1
g72R	47835	47569	267	88	no	6QX5_G, Portal protein { <i>Escherichia</i> phage T7} (91.71%/0.31)	hypothetical protein HWD19_gp25 [<i>Salmonella</i> phage yarpen]	97	3e-11	40.22	YP_009857616.1
							hypothetical protein triss_27 [<i>Salmonella</i> phage triss]	97	2e-10	40.22	QIN99047.1
g73R	48157	47798	360	119	no	No significant hits	No significant similarity found				
g74R	48593	48144	450	149	yes	No significant hits	hypothetical protein BJD50_gp72 [<i>Salmonella</i> phage BP63]	95	9e-69	71.83	YP_009302991.1
							hypothetical protein vBSenM1_22 [<i>Salmonella</i> phage vB_SenM-1]	95	9e-69	70.83	QJQ40022.1
g75R	48843	48568	276	91	yes	No significant hits	No significant similarity found				
g76R	49084	48845	240	79	no	No significant hits	hypothetical protein HWC20_gp06 [<i>Salmonella</i> phage SE4]	100	8e-39	77.22	YP_009845586.1
							hypothetical protein HOV36_gp52 [<i>Salmonella</i> phage ZCSE2]	88	3e-32	71.43	YP_009821767.1
g77R	49534	49175	360	119	no	No significant hits	hypothetical protein HWC20_gp05 [<i>Salmonella</i> phage SE4]	100	4e-65	80.67	YP_009845585.1
							hypothetical protein [<i>Salmonella</i> virus S144]	100	1e-31	53.60	QMV47891.1
g78R	49956	49531	426	141	yes	No significant hits	hypothetical protein HWC20_gp04 [<i>Salmonella</i> phage SE4]	100	9e-56	58.87	YP_009845584.1
							hypothetical protein HOV36_gp54 [<i>Salmonella</i> phage ZCSE2]	100	2e-55	58.16	YP_009821769.1
g79	50114	50443	330	109	yes	No significant hits	hypothetical protein emhyr_24 [<i>Salmonella</i> phage emhyr]	97	6e-49	72.64	QIN98901.1
							glutaredoxin [<i>Salmonella</i> phage ZCSE2]	97	9e-47	70.75	YP_009821770.1
g80	50421	52952	2532	843	yes	6A9W_A, Primase { <i>Nitratiruptor</i> phage NrS-1} (99.88%/8.9*10 ⁻²¹)	DNA primase/helicase [<i>Salmonella</i> phage SE13]	100	0.0	79.15	YP_009845509.1
							DNA primase/helicase [<i>Salmonella</i> phage vB_SenM-1]	100	0.0	78.91	QJQ40027.1

Table S3. Primers to amplify control genes encoding toxic and non-toxic gene products.

Phage	Gene	Primer	Sequence (5' – 3')	Restriction site	Size
<i>Toxic gene products</i>					
fHy-Kpn01	<i>g10</i>	Gp10FW	GCAG <u>CGGCCG</u> CATGATTAAGTACGATGTATACAAG	<i>NotI</i>	210
		Gp10RV	GGT <u>CCATGG</u> CTACTGTGAGCATAGGCTG	<i>NcoI</i>	
	<i>g22</i>	Gp22FW	GCAG <u>CGGCCG</u> CATGATTGACAGAGAAGAGATAC	<i>NotI</i>	549
		Gp22RV	GGT <u>CCATGG</u> TTAATATGCATCACGCACC	<i>NcoI</i>	
	<i>g38</i>	Gp38FW	GCAG <u>CGGCCG</u> CATGAAGTTAAACACACTAGTAA	<i>NotI</i>	441
		Gp38RV	GGT <u>CCATGG</u> TCATTCTGACCTCACTAAATG	<i>NcoI</i>	
φR1-RT	<i>g137</i>	Gp137FW	GCAG <u>CGGCCG</u> CATGAAAATTGCTGAACTAAT	<i>NotI</i>	183
		Gp137RV	GGT <u>CCATGG</u> CTATGAGGACTTAGAAATTGT	<i>NcoI</i>	
T4	<i>regB</i>	RegBFW	GATC <u>GCGGCCG</u> CCCATGACTATCAATACAGAAG	<i>NotI</i>	491
		RegBRV	GGCC <u>GCTAGC</u> TTACCTCATTGAGTTTTAATTAC	<i>NheI</i>	
<i>Non-toxic gene products</i>					
φR1-RT	<i>g119</i>	Gp119FW	GCAG <u>CGGCCG</u> CATGAAAACGTATAAAGAATTTTTG	<i>NotI</i>	591
		Gp119RV	GGT <u>CCATGG</u> TTAACGAACGTTAGTGCCA	<i>NcoI</i>	
	<i>g121</i>	Gp121FW	GCAG <u>CGGCCG</u> CATGAAAACCTATAATGAATTTATC	<i>NotI</i>	273
		Gp121RV	GGT <u>CCATGG</u> TTAGGAAGCTTTTTTAAGC	<i>NcoI</i>	
	<i>g150</i>	Gp150FW	GCAG <u>CGGCCG</u> CATGATTAAAGTTAATGAGC	<i>NotI</i>	1728
		Gp150RV	GGT <u>GCTAGC</u> TATCCAATATCAATTCGTGAA	<i>NheI</i>	
	<i>g178</i>	Gp178FW	GCAG <u>CGGCCG</u> CATGAGCAATATTAACCAGC	<i>NotI</i>	1275
		Gp178RV	GGT <u>CCATGG</u> TATCCTGCTATTAGTTTAGG	<i>NcoI</i>	
	<i>g246</i>	Gp246FW	GCAG <u>CGGCCG</u> CATGTCTTTAAATGAAATG	<i>NotI</i>	939
		Gp246RV	GGT <u>CCATGG</u> TAAAAATCATTTGCATG	<i>NcoI</i>	

Table S4. Primers used to amplify the HPUF genes of fHy-Eco03. Sequences are presented in 5' to 3' direction and restriction sites (*NcoI*, *NheI*) are underlined. Sizes of the genes in basepairs (bp) are given including the start and the stop codons.

Gene	Forward primer	Reverse primer	Size (bp)
<i>g01</i>	GCAGCGGCCGCATGAGCTATACTGACCAACA	GGTCCATGGTTACAGTTGACCTCTCAGG	198
<i>g03</i>	GCAGCGGCCCGTGTATTGGCTTAATCCTGG	GGTCCATGGTTAGTCTTGTTTCAACTCATCG	186
<i>g04</i>	GCAGCGGCCGCATGTTTAAATTCTTACAACGTAACC	GGTCCATGGTCACAAGTAATCTCCTTCAAC	180
<i>g05</i>	GCAGCGGCCGCATGTTTTCTGAGGAGCAACT	GGTCCATGGTTATTTGCTCCTTACTCCAAGT	246
<i>g06</i>	GCAGCGGCCGCATGAAAAAGCATGTTGTTGAG	GGTCCATGGTCAGAAAACATAAGACAACACC	120
<i>g08</i>	GCAGCGGCCGCATGCCATTAATCAAAGTTACAG	GGTCCATGGTCAAGAAATAATTACCTCAGTAAC	201
<i>g09</i>	GCAGCGGCCGCATGGTAACATTCACAACATATCC	GGTCCATGGTTAAACCTCAAATTTTGATTCGTC	228
<i>g10</i>	GCAGCGGCCGCATGATGAAAGAATTGACATTGAC	GGTCCATGGCTATTGTAACCTTGTACTAATTGA	177
<i>g11</i>	GCAGCGGCCGCATGCATACTCAATTAATCATCTGG	GGTCCATGGTCATTCTCCTGAGTTAAATTGTG	171
<i>g13</i>	GCAGCGGCCGCATGACTGTAGTTGCACCT	GGTCCATGGTCATTTTACCTTAAATTTAAGCCA	237
<i>g15</i>	GCAGCGGCCGCATGAACATCATAGTTGAGGGA	GGTCCATGGTCATTTAGTCCACTCCTTATCG	507
<i>g17</i>	GCAGCGGCCGCATGGACTTATTCGACTTAGTTGA	GGTCCATGGTTATGCCATTTCTACCAGCA	291
<i>g18</i>	GCAGCGGCCGCATGGCACGAAAACCACAA	GGTCCATGGTCAAGATTCATTATTATCCTCGTCC	324
<i>g23</i>	GCAGCGGCCGCATGGGTGGCGCAAGAATC	GGTCCATGGCTAATTTTGTACAAGTAAACCAGAG	369
<i>g26</i>	GCAGCGGCCGCATGGGCCTGTTGAAGATTAC	GGTCCATGGTCAGGTAATTGTAAAAGCCGG	402

<i>g40</i>	GCAGCGGCCGCATGAGAACCAGACGACTTGA	GGTCCATGGCTACCCGCAACAGTGTC	360
<i>g48</i>	GCAGCGGCCGCTTGACCTCATGCAGTAAAAAG	GGTCCATGGTTAATTGGCCCCTTCACT	216
<i>g50</i>	GCAGCGGCCGCATGATTGAATCTACACGCAC	GGTCCATGGTTAGCGCTTTCCTTTCTTTT	225
<i>g51</i>	GCAGCGGCCGCATGAGAAAGTTTACCCTGGA	GGTCCATGGTCAATCATATTTCTTAGCCAAGT	156
<i>g52</i>	GCAGCGGCCGCGTGTCTAATATAGGTTTAGAGTACC	GGTCCATGGTCATTTGCATTCTTCCTCG	276
<i>g58</i>	GCAGCGGCCGCTTGATATTCATGCTTCGCTC	GGTCCATGGTCATGCAACACCATCCTTGTT	147
<i>g59</i>	GCAGCGGCCGCTTGAATGTAGCATTAAGTGTGGC	GGTCCATGGTCAGCGAGCAGACTCGAAC	126
<i>g61</i>	GCAGCGGCCGCTTGGGCAACCGAGAACAG	GGTCCATGGCTAAAAAGGCACTTTATAGTCAAAAG	216
<i>g64</i>	GCAGCGGCCGCATGTTTGTGCCGATTCCG	GGTCCATGGTCAATTTCGCTTCTTCAGCG	549
<i>g65</i>	GCAGCGGCCGCATGCAGAACCTTTGGGATA	GGTCCATGGTCAGCTTGTGTAGCCGG	450
<i>g68</i>	GCAGCGGCCGCATGTATACGTTCAATCAATTCAAA	GGTCCATGGTTAATACCCCTCTGTTAACATC	231
<i>g69</i>	GCAGCGGCCGCATGGAAGCCACCAAGTCT	GGTCCATGGTTATTTATACACTTCAATGGTACAC	423
<i>g71</i>	GCAGCGGCCGCATGAGTGAGAACTTCGAGAAG	GGTCCATGGTTATTTTAGTGCCCCTTCGCA	300
<i>g72</i>	GCAGCGGCCGCATGAAAGCGGTCAACAAGAA	GGTCCATGGTCATGGCAGTTCTACCTGTAC	267
<i>g73</i>	GCAGCGGCCGCATGGCGACGAGTAAAAACG	GGTCCATGGTTATCTTCCTGACTGTACTTCTTG	360
<i>g76</i>	GCAGCGGCCGCATGAAACATCAGGGCACA	GGTCCATGGTTATCTCCAGCGTGCATG	240
<i>g77</i>	GCAGCGGCCGCATGAAAACAATGTTTGTATACAAG	GGTCCATGGTCACAGGCCTAACACCTC	360

Table S5. Workflow for the NGS assay bioinformatics. The raw NGS read data was analyzed using the Puhti computer environment at CSC (the Finnish Centre for Scientific Computing) following the protocol outlined below.

Step	Description	Command or script
1	The sequences over the ligation joints in the predicted plasmids on both sides of the HPUF gene fragment were determined <i>in silico</i> as described for table S8 and illustrated in figure 1b. A text document containing each sequence on its own line was prepared and saved under the name list.txt . This file was uploaded by WinSCP to the Puhti directory containing the NGS raw data files.	
2	Bio tools were activated using the commands	<pre>\$ module load biokit and \$ module load velvet</pre>
3	The compressed fastq.gz NGS sequence read files were uncompressed with the gunzip command	<pre>\$ gunzip file_name.gz</pre>
4	The paired end fastq-files were interleaved to a single file using the shufflesequences command.	<pre>\$ shuffleSequences_fastq.pl read_file_1.fq read_file_2.fq file_name.fastq</pre>
5	Alignments were run as a batch job by first editing the bash script text file template (file_name.sh) to include the required file names and paths	<pre>#!/bin/bash -l #SBATCH -o std1.out #SBATCH -e std1.err #SBATCH -p small #SBATCH --account=skurnik #SBATCH --ntasks=1 #SBATCH --cpus-per-task=1 #SBATCH --nodes=1 #SBATCH -t 48:00:00 #SBATCH --mem=128000</pre>


```

module load biokit
#change directory to the one where you
have the data
cd /file_path

# insert the file name for your sequence
segment text file
a=1
for pat in $(cat
sequencelist_file_name.txt)
do
#insert the fastq filename of your
interleaved paired end reads
fuzznuc -pattern "$pat"
file_name.fastq -rformat excel -filter |
awk '{ if ( $1 != "SeqName") print $1 }' |
sort | uniq > name_${a}
(( a = a + 1 ))
done

```

6	The bash file can be edited using Nano (version 2.3.1) with the nano command	\$ nano file_name.sh
7	To save the changes and to exit Nano the CTRL+O and CTRL+X commands were used	
8	The batch job was submitted by using the sbatch command	\$ sbatch file_name.sh
9	The batch job results in a number of files with a name containing a running number corresponding to each sequence line in the list.txt file created in point 1. Each name_N file created contains the information of the number of reads containing the searched sequence	

Table S6. LC-MS/MS analysis of the proteome of the fHy-Eco03 virion particles with the hits sorted by their decreasing score values. Note that the values in last three columns represent the encoded amino acids within the full open reading frame inside which the predicted gene resides.

Gene	Predicted function of the gene product	Gene location	Score	Coverage	# Peptides	# PSMs	# AAs	MW [kDa]	calc. pI
product									
Gp35	Tape measure protein	23078..24769	289,05	60,35	29	79	565	61,2	9,06
Gp29	No significant hits	15169..18132	228,44	52,48	40	66	987	111,5	5,33
Gp22	Major capsid protein	11285..12253	206,47	70,06	18	53	324	34,6	5,07
Gp30	Tail tube protein	18199..18630	144,07	85,71	14	47	154	16,4	5,02
Gp41	Baseplate wedge protein	27957..29114	168,29	59,08	19	46	391	42,6	4,77
Gp36	No significant hits	24766..25665	112,65	22,67	6	37	300	32,5	5,07
Gp70	No significant hits	(46148..47200)c	112,32	53,52	17	35	355	40,1	8,65
Gp80	DNA primase / helicase	50421..52952	90,97	36,57	23	30	845	96,4	6,61
Gp67	Putative NTP hydrolase	(44495..45388)c	76,39	41,75	12	23	297	33,6	5,88
Gp28	No significant hits	14556..15176	79,88	68,57	8	20	210	21,8	5,15
Gp27	Tail-to-head joining protein	14020..14553	56,33	42,50	10	19	200	22,5	6,54
Gp55	Thymidylate synthase	(36983..37912)c	67,18	58,59	11	18	326	36,9	5,82
Gp31	Outer capsid protein	18670..20787	42,43	14,41	11	17	708	73,8	4,40
Gp43	Tail fiber protein	29750..31099	56,48	50,78	12	17	451	48,6	6,89
Gp20	Portal protein	8842..10389	58,74	40,61	10	16	522	58,6	5,03
Gp32	Tail sheath protein	20874..22007	48,90	32,22	9	14	388	42,8	5,10
Gp37	No significant hits	25662..26054	46,54	55,15	5	14	136	14,9	4,88
Gp14	Dihydrofolate reductase	4697..5368	37,98	47,79	7	12	226	25,4	6,40
Gp39	No significant hits	26945..27586	47,75	44,44	5	12	216	23,1	4,84
Gp62	ATP dependent helicase	(40268..41941)c	37,36	27,03	9	12	566	63,0	8,51
Gp21	Scaffold protein	10566..11282	45,45	33,75	4	11	240	26,8	6,20
Gp16	No significant hits	5868..6590	32,98	35,42	5	8	240	27,1	5,64
Gp19	Large terminase subunit	7389..8840	25,81	22,94	6	8	497	56,2	7,39

Gp25	No significant hits	13226..13693	21,57	35,54	5	8	166	19,0	10,18
Gp60	No significant hits	(40268..41941)c	29,40	44,91	6	8	167	18,1	8,48
Gp12	No significant hits	3967..4461	23,45	46,95	3	7	164	19,1	5,07
Gp33	DUF3277 domain-containing protein	22019..22450	18,90	43,75	6	7	144	15,8	5,22
Gp38	Baseplate protein	25996..26943	24,50	24,07	5	7	324	35,7	8,19
Gp53	DNA polymerase	(33986..35953)c	25,84	13,22	5	7	658	73,5	6,70
Gp42	No significant hits	29107..29757	22,45	33,49	5	6	218	24,5	6,18
Gp66	No significant hits	(43866..44393)c	15,56	46,93	3	6	179	19,0	5,05
Gp34	No significant hits	22469..22927	20,70	24,20	3	5	157	16,7	6,57
Gp56	NTP hydrolase	(38501..38899)c	16,88	21,95	4	5	205	23,6	5,25
Gp74	No significant hits	(48144..48593)c	13,78	31,33	3	5	150	16,8	8,90
Gp24	Head-to-tail connector	12730..13224	12,48	26,49	3	4	185	20,8	5,74
Gp45	No significant hits	31727..31972	12,06	12,06	2	4	141	15,6	8,94
Gp63	Exonuclease	(42005..42877)c	13,99	16,06	3	4	249	29,0	7,43
Gp57	No significant hits	(38501..38899)c	11,22	23,60	3	4	161	19,3	9,03
Gp53.1	No significant hits	(35458..35916)c	12,27	30,67	4	4	163	18,3	11,17
Gp78	No significant hits	(49531..49956)c	16,04	28,48	2	4	151	17,4	5,34
Gp75	No significant hits	(48568..48843)c	12,12	60,38	3	4	106	12,2	4,88
Gp79	No significant hits	50114..50443	6,22	28,80	3	3	125	14,3	9,26
Gp02	Holliday-junction resolvase	311..736	3,32	21,97	2	2	173	19,5	8,62
Gp44	Til fiber assembly protein	31096..31629	5,71	16,13	2	2	186	20,6	5,08
Gp46	Lysozyme	31947..32510	3,83	16,75	2	2	191	21,2	9,00
Gp47	Spanin	32480..32830	7,81	18,75	2	2	128	13,9	8,98
Gp49	No significant hits	(33042..33320)c	6,10	17,17	2	2	99	11,0	7,28
Gp07	No significant hits	(1550..2122)c	4,31	12,32	2	2	203	22,5	5,27

Table S7. Results from plating assay of control genes.

Type of gene	Gene	Experiment 1			Experiment 2			Average	SD
		CFU1	CFU2	CFU3	CFU1	CFU2	CFU3		
Toxic	<i>g137</i>	85	103	125	114	122	104	109	13
	<i>g38</i>	65	57	93	78	54	77	71	13
	<i>g22</i>	16	17	16	12	10	19	15	3
	<i>g10</i>	8	12	15	4	5	8	9	4
	<i>regB</i>	4	5	5	8	7	12	7	3
Non-toxic	<i>g178</i>	760	624	788	451	391	447	577	157
	<i>g119</i>	350	401	352	742	797	794	573	206
	<i>g121</i>	366	335	320	375	378	312	348	26
	<i>g246</i>	247	278	246	350	376	332	305	51
	<i>g150</i>	199	200	189	356	312	376	272	78

Table S8. Results of the NGS assay with control genes. For each gene, two pairs of complementary sequences over the ligation joints were determined *in silico* containing 15-25 nt of the gene and plasmid specific sequences (the latter underlined). The joints named as *gx_atg* and *gx_stop* are sequences specific for the forward strand containing the start or stop codons of the inserted genes, while sequences named *gx_atgrev* and *gx_stoprev* are the corresponding complementary sequences from the reverse strand. Elp1 and Elp2 stand for plasmid DNA isolated from replicate transformations. The reads containing these joint sequences were extracted from the NGS sequence read files as described in Table S6. The joint-read totals for each gene and their percentages of total joint-reads were calculated. These were then used to calculate the plasmid DNA ligation mixture abundance ratios.

Gene	Joint	Joint sequence	Pooled ligation mixture			Pooled plasmid DNA				Plasmid/ligation ratio
			<i>N</i> joint-reads	Gene total	Gene %	<i>N</i> joint-reads Elp1	<i>N</i> joint-reads Elp2	Gene total	Gene %	
<i>g10</i>	<i>g10_atg</i>	<u>CCAAGCTTGCATGCCTGCAGGCGGCCGCATGATTAAGTACGATGT</u>	36			11	29			
	<i>g10_atgrev</i>	TATACATCGTACTTAATCAT <u>GCGGCCGCCTGCAGGCATGCAAGCT</u>	3			10	13			
	<i>g10_stop</i>	CAGCCTATGCTCACAGTAG <u>CCATGGAGCTAGCTCTAGAGGATCCC</u>	16			148	124			
	<i>g10_stoprev</i>	<u>GGGATCCTCTAGAGCTAGCTCCATGGCTACTGTGAGCATAGGCTG</u>	53	108	0,86	152	143	630	0,13	0,15
<i>g22</i>	<i>g22_atg</i>	<u>CCAAGCTTGCATGCCTGCAGGCGGCCGCATGATTGACAGAGAAGA</u>	193			22	13			
	<i>g22_atgrev</i>	TCTCTTCTGTGTCAATCAT <u>GCGGCCGCCTGCAGGCATGCAAGCTT</u>	0			32	7			
	<i>g22_stop</i>	GGTGCCTGATGCATATTAAC <u>CATGGAGCTAGCTCTAGAGGATCCC</u>	5			40	94			
	<i>g22_stoprev</i>	<u>GGGATCCTCTAGAGCTAGCTCCATGGTTAATATGCATCACGCACC</u>	193	391	3,10	21	94	323	0,07	0,02
<i>g38</i>	<i>g38_atg</i>	<u>CCAAGCTTGCATGCCTGCAGGCGGCCGCATGAAGTTAAACACACTAG</u>	1229			17191	16249			
	<i>g38_atgrev</i>	TTACTAGTGTGTTTTAACTTCAT <u>GCGGCCGCCTGCAGGCATGCAAGCT</u>	3			14108	13946			
	<i>g38_stop</i>	GGGATCCTCTAGAGCTAGCTCCATGGTCACTTCTGACCTCACTAAATG	614			17730	16471			
	<i>g38_stoprev</i>	CATTTAGTGAGGTCAGAATGA <u>CCATGGAGCTAGCTCTAGAGGATCCC</u>	23	1869	14,84	17760	16749	130204	26,83	1,81
<i>g119</i>	<i>g119_atg</i>	<u>CCAAGCTTGCATGCCTGCAGGCGGCCGCATGAAAACGTATAAAGAATT</u>	877			11986	11896			
	<i>g119_atgrev</i>	AATTCTTTATACGTTTTTCAT <u>GCGGCCGCCTGCAGGCATGCAAGCTTGG</u>	4			11539	11460			
	<i>g119_stop</i>	TGGCACTAACGTTTCGTTAA <u>CCATGGAGCTAGCTCTAGAGGATCCC</u>	26			12216	12346			
	<i>g119_stoprev</i>	<u>GGGATCCTCTAGAGCTAGCTCCATGGTTAACGAACGTTAGTGCCA</u>	567	1474	11,70	11503	11787	94733	19,52	1,67
<i>g121</i>	<i>g121_atg</i>	<u>CCAAGCTTGCATGCCTGCAGGCGGCCGCATGAAAACCTATAATGAATT</u>	156			9467	9211			
	<i>g121_atgrev</i>	AATTCATTATAGGTTTTTCAT <u>GCGGCCGCCTGCAGGCATGCAAGCTTGG</u>	2			10055	10311			
	<i>g121_stop</i>	GCTTAAAAAAGCTTCCTAA <u>CCATGGAGCTAGCTCTAGAGGATCCC</u>	24			9823	9688			
	<i>g121_stoprev</i>	<u>GGGATCCTCTAGAGCTAGCTCCATGGTTAGGAAGCTTTTTTAAGC</u>	170	352	2,79	9087	8714	76356	15,74	5,63
<i>g137</i>	<i>g137_atg</i>	<u>CCAAGCTTGCATGCCTGCAGGCGGCCGCATGAAAATGCTGAACATAAT</u>	16			7000	7362			
	<i>g137_atgrev</i>	ATTAGTTCAGCAATTTTCAT <u>GCGGCCGCCTGCAGGCATGCAAGCTTGG</u>	0			8356	8866			

	<i>g137_stop</i>	<u>ACAATTTCTAAGTCCTCATAGCCATGGAGCTAGCTCTAGAGGATCCC</u>	15			9320	9287		
	<i>g137_stoprev</i>	<u>GGGATCCTCTAGAGCTAGCTCCATGGCTATGAGGACTTAGAAATTGT</u>	30	61	0,48	6716	6924	63831	13,15
<i>g150</i>	<i>g150_atg</i>	<u>CCAAGCTTGCATGCCTGCAGGCGGCCGCATGATTAAAGTTAATGAGC</u>	1641			2398	2625		
	<i>g150_atgrev</i>	<u>GCTCATTAACTTTAATCATGCGGCCGCCTGCAGGCATGCAAGCTTGG</u>	3			2320	2489		
	<i>g150_stop</i>	<u>CACGAATTGATATTGGATAGGCTAGCTCTAGAGGATCCCCGGGTAC</u>	17			2162	2245		
	<i>g150_stoprev</i>	<u>GTACCGGGGATCCTCTAGAGCTAGCCTATCCAATATCAATTTCGTG</u>	1032	2693	21,38	1950	2288	18477	3,81
<i>g178</i>	<i>g178_atg</i>	<u>CCAAGCTTGCATGCCTGCAGGCGGCCGCATGAGCAATATTAACCAGC</u>	1387			7335	6952		
	<i>g178_atgrev</i>	<u>GCTGGTTAATATTGCTCATGCGGCCGCCTGCAGGCATGCAAGCTTGG</u>	0			6305	6202		
	<i>g178_stop</i>	<u>AAACTAATAGCAGGATAACCATGGAGCTAGCTCTAGAGGATCCC</u>	9			6111	5619		
	<i>g178_stoprev</i>	<u>GGGATCCTCTAGAGCTAGCTCCATGGTTATCCTGCTATTAGTTT</u>	1342	2738	21,74	6462	6341	51327	10,58
<i>g246</i>	<i>g246_atg</i>	<u>CCAAGCTTGCATGCCTGCAGGCGGCCGCATGTCTTTAAATGAAATG</u>	1487			6475	6464		
	<i>g246_atgrev</i>	<u>CATTTTCATTTAAAGACATGCGGCCGCCTGCAGGCATGCAAGCTTGG</u>	9			6219	6227		
	<i>g246_stop</i>	<u>CATGCAAAATGATTTTAAACCATGGAGCTAGCTCTAGAGGATCCC</u>	19			6789	6478		
	<i>g246_stoprev</i>	<u>GGGATCCTCTAGAGCTAGCTCCATGGTTAAAAATCATTTCATG</u>	1396	2911	23,11	5442	5254	49348	10,17
<i>regB</i>	<i>regB_atg</i>	<u>CCAAGCTTGCATGCCTGCAGGCGGCCGCATGACTATCAATACAG</u>	0			0	0		
	<i>regB_atgrev</i>	<u>CTGTATTGATAGTCATGCGGCCGCCTGCAGGCATGCAAGCTTGG</u>	0			0	0		
	<i>regB_stop</i>	<u>ATTAAAACTCAATGAGGTAAGGCTAGCTCTAGAGGATCCCC</u>	0			0	0		
	<i>regB_stoprev</i>	<u>GGGGATCCTCTAGAGCTAGCCTTACCTCATTGAGTTTAAAT</u>	0	0	0,00	0	0	0	0,00
Total reads				12597	100%	485229			
						100%			

Table S9. Plating assay results of 23 HPUF-encoding genes of fHy-Eco03.

	Control genes					Ligation 1					Ligation 2					Fraction of <i>g178</i>			
	CFU1	CFU2	CFU3	Average	SD	CFU1	CFU2	CFU3	Average	SD	CFU1	CFU2	CFU3	Average	SD	Lig 1	Lig 2	Average	SD
<i>regB</i>	11	9	17	12	3														
<i>g178</i>	391	451	447	430	27														
<i>g01</i>						353	321	419	364	41	186	303	175	221	58	0.847	0.515	0.681	0.166
<i>g03</i>						448	446	419	438	13	565	362	349	425	99	1.018	0.989	1.003	0.014
<i>g04</i>						263	312	247	274	28	278	324	311	304	19	0.637	0.708	0.672	0.035
<i>g05</i>						47	80	81	69	16	81	83	52	72	14	0.161	0.167	0.164	0.003
<i>g06</i>						400	389	343	377	25	225	222	185	211	18	0.878	0.49	0.684	0.194
<i>g08</i>						309	364	344	339	23	243	181	258	227	33	0.788	0.529	0.659	0.130
<i>g09</i>						432	400	423	418	13	312	290	277	293	14	0.973	0.681	0.827	0.146
<i>g10</i>						314	308	214	279	46	97	128	133	119	16	0.648	0.278	0.463	0.185
<i>regB</i>	9	12	9	10	1														
<i>g178</i>	417	282	314	338	58														
<i>g11</i>						399	380	388	389	8	173	149	337	220	84	1.151	0.65	0.900	0.250
<i>g13</i>						504	625	501	543	58	205	314	242	254	45	1.607	0.75	1.179	0.429
<i>g15</i>						482	479	513	491	15	244	282	239	255	19	1.454	0.754	1.104	0.350
<i>g17</i>						382	328	396	369	29						1.091		0.645	0.223
<i>regB</i>	7	8	12	9	2														
<i>g178</i>	397	415	413	408	8														
<i>g23</i>						196	191	205	197	6						0.484		0.682	0.199
<i>g48</i>						620	680	666	655	26						1.606		1.711	0.105
<i>g50</i>						449	370	382	400	35						0.981		0.747	0.234
<i>g51</i>						242	279	239	253	18	100	125	94	106	13	0.621	0.261	0.441	0.180
<i>regB</i>	11	10	10	10	0														
<i>g178</i>	742	597	794	711	83														
<i>g52</i>						486	348	331	388	69	335	319	252	302	36	0.546	0.425	0.485	0.061

[illegible]

Table S10. NGS screening of phage fHy-Eco03 HPUF encoding genes. For each gene, two pairs of complementary sequences over the ligation joints were determined *in silico* containing 15-25 nt of the gene and plasmid specific sequences (the latter underlined). The joints named as *gx_atg* and *gx_stop* are sequences specific for the forward strand containing the start or stop codons of the inserted genes, while sequences named *gx_atgrev* and *gx_stoprev* are the corresponding complementary sequences from the reverse strand. Ligation Mix 1 and Mix 2 represent replicate ligation mixtures and Mix(1/2). Elp(1/2) represent replicate electroporations of the ligation mixes. The reads containing these joint sequences were extracted from the NGS sequence read files as described in Table S6. The joint-read totals for each gene and their percentages of total joint-reads were calculated. These were then used to calculate the plasmid DNA ligation mixture abundance ratios.

Gene	Joint	Joint sequence	Pooled ligations				Plasmid DNA from pooled transformants						% ratios
			Mix 1	Mix 2	N per gene	% per gene	Mix1Elp1	Mix1Elp2	Mix2Elp1	Mix2Elp2	N per gene	% per gene	
NGS experiment 1													
g01	g01_atg	<u>CCAAGCTTGCATGCCTGCAGGCGGCCGCATGAGCTATACTGACCAACA</u>	95	55			3562	3244	3162	2449			
	g01_atgrev	<u>TGTTGGTCAGTATAGCTCATGCGGCCGCCTGCAGGCATGCAAGCTTGG</u>	0	2			3930	3443	3243	2451			
	g01_stop	<u>CCTGAGAGGTCAACTGTAACTAGGAGCTAGCTCTAGAGGATCCC</u>	24	27			3704	3186	3293	2471			
	g01_stoprev	<u>GGGATCCTCTAGAGCTAGTCCATGGTTACAGTTGACCTCTCAGG</u>	126	105	434	1,31	3665	3312	3257	2484	50856	3,52	2,69
g03	g03_atg	<u>CCAAGCTTGCATGCCTGCAGGCGGCCGCTGATTGGCTTAATCCTTGG</u>	104	81			5920	5084	4659	3780			
	g03_atgrev	<u>CCAGGATTAAGCCAATCACGCGGCCGCTGCAGGCATGCAAGCTTGG</u>	2	2			5973	4799	4627	3435			
	g03_stop	<u>CGATGAGTTGAAACAAGACTAACCATGGAGCTAGCTCTAGAGGATCCC</u>	14	24			5938	5086	4863	3754			
	g03_stoprev	<u>GGGATCCTCTAGAGCTAGTCCATGGTTAGTCTTGTTCACCTCATCG</u>	103	88	418	1,26	5770	4927	4656	3613	76884	5,33	4,22
g04	g04_atg	<u>CCAAGCTTGCATGCCTGCAGGCGGCCGCATGTTTAAATTCCTACAACGTAACC</u>	58	53			3239	2652	2607	2197			
	g04_atgrev	<u>GGTTACGTTGTAAGAATTTAAACATGCGGCCGCCTGCAGGCATGCAAGCTTGG</u>	4	2			3014	2477	2352	1885			
	g04_stop	<u>GTTGAAGGAGATTACTTGTGACCATGGAGCTAGCTCTAGAGGATCCC</u>	26	18			3240	2553	2760	2184			
	g04_stoprev	<u>GGGATCCTCTAGAGCTAGTCCATGGTCACAAGTAATCTCCTTCAAC</u>	77	56	294	0,89	3347	2763	2729	2243	42242	2,93	3,30
g05	g05_atg	<u>CCAAGCTTGCATGCCTGCAGGCGGCCGCATGTTTCTGAGGAGCAACT</u>	283	178			0	3	4	0			
	g05_atgrev	<u>AGTTGCTCCTCAGAAAACATGCGGCCGCCTGCAGGCATGCAAGCTTGG</u>	2	1			0	0	0	0			
	g05_stop	<u>ACTTGAGTAAGGAGCAAATAACCATGGAGCTAGCTCTAGAGGATCCC</u>	42	25			21	10	12	8			
	g05_stoprev	<u>GGGATCCTCTAGAGCTAGTCCATGGTTATTTGCTCCTTACTCCAAGT</u>	335	248	1114	3,36	33	38	5	12	146	0,01	0,003
g06	g06_atg	<u>CCAAGCTTGCATGCCTGCAGGCGGCCGCATGAAAAGCATGTTGTTGAG</u>	27	26			3666	3012	3342	2638			
	g06_atgrev	<u>CTCAACAACATGCTTTTTCATGCGGCCGCCTGCAGGCATGCAAGCTTGG</u>	8	11			3637	3076	3264	2626			
	g06_stop	<u>GGTGTGTCTTATGTTTCTGACCATGGAGCTAGCTCTAGAGGATCCC</u>	156	175			3611	3256	3578	2895			
	g06_stoprev	<u>GGGATCCTCTAGAGCTAGTCCATGGTCAGAAAACATAAGACAACACC</u>	46	36	485	1,46	3562	3072	3373	2691	51299	3,56	2,43
g08	g08_atg	<u>CCAAGCTTGCATGCCTGCAGGCGGCCGCATGCCATTAATCAAAGTTACAG</u>	80	46			4148	3417	3805	3052			
	g08_atgrev	<u>CTGTAACCTTGTATTAAATGGCATGCGGCCGCCTGCAGGCATGCAAGCTTGG</u>	3	1			3528	3075	3626	2631			
	g08_stop	<u>GTTACTGAGGTAATTTCTTGAACCATGGAGCTAGCTCTAGAGGATCCC</u>	41	26			4010	3509	4102	3220			
	g08_stoprev	<u>GGGATCCTCTAGAGCTAGTCCATGGTCAAGAAATAATTACCTCAGTAAC</u>	121	106	424	1,28	3581	3219	3749	2867	55539	3,85	3,01
g09	g09_atg	<u>CCAAGCTTGCATGCCTGCAGGCGGCCGCATGGTAACATTCAACAATATCC</u>	251	142			5079	4545	3602	3039			
	g09_atgrev	<u>GGATAGTTGTGAATGTTTACCATGCGGCCGCCTGCAGGCATGCAAGCTTGG</u>	10	3			5054	4295	3427	2766			
	g09_stop	<u>GACGAATCAAAATTTGAGGTTTAAACCATGGAGCTAGCTCTAGAGGATCCC</u>	35	29			4920	4132	3430	2741			
	g09_stoprev	<u>GGGATCCTCTAGAGCTAGTCCATGGTTAAACCTCAAATTTGATTGCTC</u>	281	178	929	2,80	4617	4191	3383	2828	62049	4,30	1,53
g10	g10_atg	<u>CCAAGCTTGCATGCCTGCAGGCGGCCGCATGATGAAAGAATTGACATTGAC</u>	80	28			2089	1723	1148	916			
	g10_atgrev	<u>GTCAATGTCAATTCTTTCATCATGCGGCCGCCTGCAGGCATGCAAGCTTGG</u>	5	60			2152	1714	1290	943			
	g10_stop	<u>TCAAATTAGTACAAAGTTACAATAGCCATGGAGCTAGCTCTAGAGGATCCC</u>	23	11			2273	1595	1317	985			
	g10_stoprev	<u>GGGATCCTCTAGAGCTAGTCCATGGCTATTGTAACCTTTGTACTAATTTGA</u>	92	60	359	1,08	1807	1549	1120	846	23467	1,63	1,50
g11	g11_atg	<u>CCAAGCTTGCATGCCTGCAGGCGGCCGCATGCATACTCAATTAATCATCTGG</u>	71	67			2949	2514	2725	2207			

	<i>g11_atgrev</i>	<u>CCAGATGATTAATTAGTATGCATGCGGCCGCTGCAGGCATGCAAGCTTGG</u>	4	5			2945	2530	2821	2149			
	<i>g11_stop</i>	<u>CACAATTTAACTCAGGAGAATGACCATGGAGCTAGCTCTAGAGGATCCC</u>	29	20			3157	2815	3035	2447			
	<i>g11_stoprev</i>	<u>GGGATCCTCTAGAGCTAGCTCCATGGTCATTCTCCTGAGTTAAATTGTG</u>	90	68	354	1,07	3006	2491	2800	2282	42873	2,97	2,78
<i>g13</i>	<i>g13_atg</i>	<u>CCAAGCTTGCATGCCTGCAGCGCGCCGCATGACTGTAGTTGCACCT</u>	183	250			6229	5350	7414	5717			
	<i>g13_atgrev</i>	<u>AGGTGCAACTACAGTCATGCGGCCGCTGCAGGCATGCAAGCTTGG</u>	2	7			4250	3905	5337	3939			
	<i>g13_stop</i>	<u>TGGCTTAAATTTAAGGTAAATGACCATGGAGCTAGCTCTAGAGGATCCC</u>	12	21			4731	4310	6048	4987			
	<i>g13_stoprev</i>	<u>GGGATCCTCTAGAGCTAGCTCCATGGTCATTTTACCTTAAATTTAAGCCA</u>	224	251	950	2,87	4921	4266	6056	4915	82375	5,71	1,99
<i>g15</i>	<i>g15_atg</i>	<u>CCAAGCTTGCATGCCTGCAGCGCGCCGCATGAACATCATAGTTGAGGGA</u>	820	595			6502	5798	3773	3065			
	<i>g15_atgrev</i>	<u>TCCCTCAACTATGATGTTTCATGCGGCCGCTGCAGGCATGCAAGCTTGG</u>	3	5			6322	5494	3775	2777			
	<i>g15_stop</i>	<u>CGATAAGGAGTGGACTAAATGACCATGGAGCTAGCTCTAGAGGATCCC</u>	42	29			6070	5806	3783	3228			
	<i>g15_stoprev</i>	<u>GGGATCCTCTAGAGCTAGCTCCATGGTCATTTAGTCCACTCCTTATCG</u>	881	488	2863	8,64	6101	5753	3711	3146	75104	5,21	0,60
<i>g17</i>	<i>g17_atg</i>	<u>CCAAGCTTGCATGCCTGCAGCGCGCCGCATGGACTTATTGCACTTAGTTGA</u>	639	455			6207	5269	4906	4122			
	<i>g17_atgrev</i>	<u>TCAACTAAGTCGAATAAGTCCATGCGGCCGCTGCAGGCATGCAAGCTTGG</u>	8	2			5034	4373	4366	3212			
	<i>g17_stop</i>	<u>TGCTGGTAGAAATGGCATAACCATGGAGCTAGCTCTAGAGGATCCC</u>	65	46			6295	5211	5472	4377			
	<i>g17_stoprev</i>	<u>GGGATCCTCTAGAGCTAGCTCCATGGTTATGCCATTTCTACCAGCA</u>	809	555	2579	7,79	5995	5116	5124	4126	79205	5,49	0,71
<i>g23</i>	<i>g23_atg</i>	<u>CCAAGCTTGCATGCCTGCAGCGCGCCGCATGGTGGCGCAAGAATC</u>	689	461			3220	2952	2936	2134			
	<i>g23_atgrev</i>	<u>GATTCTTGGGCCACCCATGCGGCCGCTGCAGGCATGCAAGCTTGG</u>	4	4			3019	2658	2716	2228			
	<i>g23_stop</i>	<u>TTTACTTGTACAAAATTAGCCATGGAGCTAGCTCTAGAG</u>	68	44			3380	3204	3293	2587			
	<i>g23_stoprev</i>	<u>CTCTAGAGCTAGCTCCATGGCTAATTTTGTACAAGTAA</u>	1522	1213	4005	12,09	3655	3378	3346	2680	47386	3,28	0,27
<i>g48</i>	<i>g48_atg</i>	<u>CCAAGCTTGCATGCCTGCAGCGCGCCGCTTGACCTCATGCAGTAAAAAG</u>	241	182			16372	14621	14149	11134			
	<i>g48_atgrev</i>	<u>CTTTTTACTGCATGAGGTCAAGCGCCGCTGCAGGCATGCAAGCTTGG</u>	6	3			14949	13359	12842	9941			
	<i>g48_stop</i>	<u>AGTGAAGGGGCCAATTAAACCATGGAGCTAGCTCTAGAGGATCCC</u>	87	79			15501	14418	14516	11213			
	<i>g48_stoprev</i>	<u>GGGATCCTCTAGAGCTAGCTCCATGGTTAATTGGCCCTTCACT</u>	282	247	1127	3,40	16598	15055	15410	11945	222023	15,39	4,52
<i>g50</i>	<i>g50_atg</i>	<u>CCAAGCTTGCATGCCTGCAGCGCGCCGCATGATTGAATCTACACGCAC</u>	244	162			5148	4293	3726	2957			
	<i>g50_atgrev</i>	<u>GTGCGGTAGATTCAATCATGCGGCCGCTGCAGGCATGCAAGCTTGG</u>	5	3			4566	3610	2966	2280			
	<i>g50_stop</i>	<u>AAAAGAAAGAAAGCGCTAAACCATGGAGCTAGCTCTAGAGGATCCC</u>	40	41			4639	3793	3206	2737			
	<i>g50_stoprev</i>	<u>GGGATCCTCTAGAGCTAGCTCCATGGTTAGCGCTTCTCTTCTTT</u>	233	155	883	2,67	4692	3994	3447	2807	58861	4,08	1,53
<i>g51</i>	<i>g51_atg</i>	<u>CCAAGCTTGCATGCCTGCAGCGCGCCGCATGAGAAAGTTTACCCTGGA</u>	50	42			5333	4457	4895	3957			
	<i>g51_atgrev</i>	<u>TCCAGGGTAACTTTTCTCATGCGGCCGCTGCAGGCATGCAAGCTTGG</u>	4	4			6204	5176	5605	4309			
	<i>g51_stop</i>	<u>ACTTGGCTAAGAAATATGATTGACCATGGAGCTAGCTCTAGAGGATCCC</u>	27	9			5561	4843	5635	4481			
	<i>g51_stoprev</i>	<u>GGGATCCTCTAGAGCTAGCTCCATGGTCAATCATATTTCTTAGCCAAGT</u>	85	80	301	0,91	5020	4254	4697	3650	78077	5,41	5,96
<i>g52</i>	<i>g52_atg</i>	<u>CCAAGCTTGCATGCCTGCAGCGCGCCGCTGTCTAATATAGGTTTAGAGTACC</u>	562	502			4254	3525	3656	2843			
	<i>g52_atgrev</i>	<u>GGTACTCTAAACCTATATTAGACACGCGGCCGCTGCAGGCATGCAAGCTTGG</u>	6	19			3384	2739	3070	2244			
	<i>g52_stop</i>	<u>CGAGGAAGAATGCAAAATGACCATGGAGCTAGCTCTAGAGGATCCC</u>	35	40			4448	3531	3906	3292			
	<i>g52_stoprev</i>	<u>GGGATCCTCTAGAGCTAGCTCCATGGTCATTTGCACTTCTCTCTCG</u>	606	465	2235	6,75	4127	3439	3598	2899	54955	3,81	0,56
<i>g64</i>	<i>g64_atg</i>	<u>CCAAGCTTGCATGCCTGCAGCGCGCCGCATGTTTGTGCCGATTCCG</u>	837	497			2055	2072	1918	1371			
	<i>g64_atgrev</i>	<u>CGGAATCGGCACAAACATGCGGCCGCTGCAGGCATGCAAGCTTGG</u>	9	4			1933	1847	1814	1396			
	<i>g64_stop</i>	<u>CGCTGAAGAAGCGAATTGACCATGGAGCTAGCTCTAGAGGATCCC</u>	30	17			2068	2012	1976	1520			
	<i>g64_stoprev</i>	<u>GGGATCCTCTAGAGCTAGCTCCATGGTCAATTCGCTTCTTCACGG</u>	746	479	2619	7,91	2320	1994	2049	1575	29920	2,07	0,26
<i>g65</i>	<i>g65_atg</i>	<u>CCAAGCTTGCATGCCTGCAGCGCGCCGCATGCAGAACCTTTGGGATA</u>	533	444			5013	4535	5843	4669			
	<i>g65_atgrev</i>	<u>TATCCCAAAGGTTCTGCATGCGGCCGCTGCAGGCATGCAAGCTTGG</u>	0	3			4243	3836	5056	3961			
	<i>g65_stop</i>	<u>CCGGCTACACAAGCTGACCATGGAGCTAGCTCTAGAGGATCCC</u>	27	30			4708	4347	5995	4460			
	<i>g65_stoprev</i>	<u>GGGATCCTCTAGAGCTAGCTCCATGGTCAGCTTGTGTAGCCGG</u>	380	369	1786	5,39	5268	4669	6475	4800	77878	5,40	1,00
<i>g68</i>	<i>g68_atg</i>	<u>CCAAGCTTGCATGCCTGCAGCGCGCCGCATGTATACGTTCAATCAAAA</u>	293	205			1348	1086	846	678			
	<i>g68_atgrev</i>	<u>TTTGAATTGATTGAACGTATACATGCGGCCGCTGCAGGCATGCAAGCTTGG</u>	8	7			1157	1034	757	553			
	<i>g68_stop</i>	<u>GATGTTAACAGAGGGGTATTAACCATGGAGCTAGCTCTAGAGGATCCC</u>	42	42			1313	1020	936	678			
	<i>g68_stoprev</i>	<u>GGGATCCTCTAGAGCTAGCTCCATGGTTAATACCCCTCTGTTAACATC</u>	336	229	1162	1,31	1305	1039	848	732	15330	1,06	0,81
<i>g69</i>	<i>g69_atg</i>	<u>CCAAGCTTGCATGCCTGCAGCGCGCCGCATGGAAGCCACCAAGTCT</u>	965	556			6676	5661	5057	4117			

	<i>g69_atgrev</i>	<u>AGACTTGGTGGCTTCCATGCGGCCGCTGCAGGCATGCAAGCTTGG</u>	3	2			6311	5534	4816	3859			
	<i>g69_stop</i>	<u>GTGTACCATTGAAGTGTATAAATAACCATGGAGCTAGCTCTAGAGGATCCC</u>	19	11			5559	4802	4518	3431			
	<i>g69_stoprev</i>	<u>GGGATCCTCTAGAGCTAGCTCCATGGTTATTATACACTTCAATGGTACAC</u>	687	467	2710	8,18	6244	5494	5048	4024	81151	5,62	0,69
<i>g76</i>	<i>g76_atg</i>	<u>AGCTTGCATGCCTGCAGGCGGCCGCATGAAACATCAGGGC</u>	355	302			5815	4711	4355	3411			
	<i>g76_atgrev</i>	<u>GCCCTGATGTTTCATGCGGCCGCTGCAGGCATGCAAGCT</u>	1	2			4961	4305	3949	3037			
	<i>g76_stop</i>	<u>CATGCACGCTGGAGATAACCATGGAGCTAGCTCTAGAGGATCCC</u>	40	26			4816	3867	3986	2842			
	<i>g76_stoprev</i>	<u>GGGATCCTCTAGAGCTAGCTCCATGGTTATCTCCAGCGTGCATG</u>	353	226	1305	3,94	5026	4241	3925	3084	66331	4,60	1,17
<i>g77</i>	<i>g77_atg</i>	<u>CCAAGCTTGCATGCCTGCAGGCGGCCGCATGAAACAATGTTGTATACAAG</u>	1426	870			5061	3957	5141	4112			
	<i>g77_atgrev</i>	<u>CTTGTATACAAACATTGTTTTCATGCGGCCGCTGCAGGCATGCAAGCTTGG</u>	12	6			3902	3272	4273	3045			
	<i>g77_stop</i>	<u>GAGGTGTTAGGCCTGTGACCATGGAGCTAGCTCTAGAGGATCCC</u>	49	35			4599	3988	5007	4006			
	<i>g77_stoprev</i>	<u>GGGATCCTCTAGAGCTAGCTCCATGGTCACAGGCCTAACACCTC</u>	709	681	3788	11,44	4720	4073	5396	4310	68862	4,77	0,42
Totals			33124				1442813						

NGS experiment 2

<i>g18</i>	<i>g18_atg</i>	<u>CCAAGCTTGCATGCCTGCAGGCGGCCGCATGGCACGAAAACCACAA</u>	594	568			32026	25686	27758	26039			
	<i>g18_atgrev</i>	<u>TTGTGGTTTTCGTGCCATGCGGCCGCTGCAGGCATGCAAGCTTGG</u>	4	0			24355	19722	20933	19404			
	<i>g18_stop</i>	<u>GGACGAGGATAATAATGAATCTTGACCATGGAGCTAGCTCTAGAGGATCCC</u>	16	4			23707	18651	20102	18915			
	<i>g18_stoprev</i>	<u>GGGATCCTCTAGAGCTAGCTCCATGGTCAAGATTCAATTATTATCTCGTCC</u>	415	291	1892	36,90	24199	18668	19867	19246	359278	41,23	1,12
<i>g26</i>	<i>g26_atg</i>	<u>CCAAGCTTGCATGCCTGCAGGCGGCCGCATGGGCTGTTGAAGATTAC</u>	215	206			8285	6530	7396	7384			
	<i>g26_atgrev</i>	<u>GTAATCTTCAACAGGCCCATGCGGCCGCTGCAGGCATGCAAGCTTGG</u>	2	0			7618	6473	7369	6877			
	<i>g26_stop</i>	<u>CCGGCTTTTACAATTACCTGACCATGGAGCTAGCTCTAGAGGATCCC</u>	19	18			7373	5499	6399	5986			
	<i>g26_stoprev</i>	<u>GGGATCCTCTAGAGCTAGCTCCATGGTCAGGTAATTGTAAAGCCGG</u>	295	231	986	19,23	7943	6355	7247	6830	111564	12,80	0,67
<i>g40</i>	<i>g40_atg</i>	<u>CCAAGCTTGCATGCCTGCAGGCGGCCGCATGAGAACCAGACGACTTGA</u>	99	58			2351	2162	2007	1861			
	<i>g40_atgrev</i>	<u>TCAAGTCGCTGGTTCTCATGCGGCCGCTGCAGGCATGCAAGCTTGG</u>	1	2			1804	1743	1680	1500			
	<i>g40_stop</i>	<u>GACACTGTTGCGGGTAGCCATGGAGCTAGCTCTAGAGGATCCC</u>	10	18			2413	1935	1954	1679			
	<i>g40_stoprev</i>	<u>GGGATCCTCTAGAGCTAGCTCCATGGCTACCCCGCAACAGTGTC</u>	258	142	588	11,47	2582	2104	2045	1833	31653	3,63	0,32
<i>g58</i>	<i>g58_atg</i>	<u>CCAAGCTTGCATGCCTGCAGGCGGCCGCTTGATATTTCATGCTTCGCTC</u>	0	0			190	96	141	104			
	<i>g58_atgrev</i>	<u>GAGCGAAGCATGAATATCCAAGCGGCCGCTGCAGGCATGCAAGCTTGG</u>	0	0			188	125	171	121			
	<i>g58_stop</i>	<u>AACAAGGATGGTTGTCATGACCATGGAGCTAGCTCTAGAGGATCCC</u>	7	10			268	172	240	252			
	<i>g58_stoprev</i>	<u>GGGATCCTCTAGAGCTAGCTCCATGGTCATGCAACACCATCTCTGTT</u>	23	9	49	0,96	249	153	190	191	2851	0,33	0,34
<i>g59</i>	<i>g59_atg</i>	<u>CCAAGCTTGCATGCCTGCAGGCGGCCGCTTGAATGTAGCATTAAGTGTGGC</u>	6	4			11217	8583	9104	8817			
	<i>g59_atgrev</i>	<u>GCCACACTTAATGCTACATTCAAGCGGCCGCTGCAGGCATGCAAGCTTGG</u>	0	0			10005	8543	8755	8422			
	<i>g59_stop</i>	<u>GTTTCGAGTCTGCTCGCTGACCATGGAGCTAGCTCTAGAGGATCCC</u>	13	10			11608	8909	9883	9073			
	<i>g59_stoprev</i>	<u>GGGATCCTCTAGAGCTAGCTCCATGGTCAGCGAGCAGACTCGAAC</u>	33	55	121	2,36	11051	8558	9485	8459	150472	17,27	7,32
<i>g61</i>	<i>g61_atg</i>	<u>CCAAGCTTGCATGCCTGCAGGCGGCCGCTTGGGCAACCAGAGACAG</u>	30	29			7293	6429	4773	4955			
	<i>g61_atgrev</i>	<u>CTGTTCTCGGTTGCCAAGCGGCCGCTGCAGGCATGCAAGCTTGG</u>	0	0			6223	5908	4512	4592			
	<i>g61_stop</i>	<u>CTTTTGACTATAAAGTGCCTTTTATAGCCATGGAGCTAGCTCTAGAGGATCCC</u>	10	1			5433	4700	3700	3641			
	<i>g61_stoprev</i>	<u>GGGATCCTCTAGAGCTAGCTCCATGGCTAAAAGGCACCTTATAGTCAAAAG</u>	44	25	139	2,71	6233	5290	4055	4224	81961	9,41	3,47
<i>g71</i>	<i>g71_atg</i>	<u>CCAAGCTTGCATGCCTGCAGGCGGCCGCATGAGTGAGAACTTCGAGAAG</u>	124	125			6723	5962	5272	5492			
	<i>g71_atgrev</i>	<u>CTTCTCGAAGTTCTCACTCATGCGGCCGCTGCAGGCATGCAAGCTTGG</u>	0	1			6472	6138	5315	5338			
	<i>g71_stop</i>	<u>TGCGAAGGGGCACTAAAATAACCATGGAGCTAGCTCTAGAGGATCCC</u>	19	13			6720	5587	4838	5011			
	<i>g71_stoprev</i>	<u>GGGATCCTCTAGAGCTAGCTCCATGGTTATTTTTAGTGCCCTTGCA</u>	182	130	594	11,59	6759	6080	5171	5267	92145	10,57	0,91
<i>g72</i>	<i>g72_atg</i>	<u>CCAAGCTTGCATGCCTGCAGGCGGCCGCATGAAAGCGGTCAACAAGAA</u>	17	25			1487	1452	1262	1081			
	<i>g72_atgrev</i>	<u>TTCTTGTGTACCGCTTTTCATGCGGCCGCTGCAGGCATGCAAGCTTGG</u>	0	2			1321	1267	1109	977			
	<i>g72_stop</i>	<u>GTACAGGTGAACCTGCCATGACCATGGAGCTAGCTCTAGAGGATCCC</u>	11	11			1585	1617	1191	1126			
	<i>g72_stoprev</i>	<u>GGGATCCTCTAGAGCTAGCTCCATGGTCATGCGAGTTCTACCTGTAC</u>	85	58	209	4,08	1620	1506	1320	1212	21133	2,43	0,59
<i>g73</i>	<i>g73_atg</i>	<u>CCAAGCTTGCATGCCTGCAGGCGGCCGCATGGCGACGAGTAAAAACG</u>	51	42			1686	1305	1175	1238			
	<i>g73_atgrev</i>	<u>CGTTTTTACTCGTCGCCATGCGGCCGCTGCAGGCATGCAAGCTTGG</u>	0	0			1444	1190	1097	1058			

<i>g73_stop</i>	CAAGAAGTACAGTCAGGAAGATAACCATGGAGCTAGCTCTAGAGGATCCC	22	17			1464	1224	1145	1083			
<i>g73_stoprev</i>	GGGATCCTCTAGAGCTAGCTCCATGGTTATCTTCCTGACTGTACTTCTTG	268	149	549	10,71	1612	1294	1187	1196	20398	2,34	0,22
Totals			5127							871455		

References

1. Mohanraj, U.; Wan, X.; Spruit, C.M.; Skurnik, M.; Pajunen, M.I. A toxicity screening approach to identify bacteriophage-encoded anti-microbial proteins. *Viruses* **2019**, *11*.

2. Zimmermann, L.; Stephens, A.; Nam, S.Z.; Rau, D.; Kubler, J.; Lozajic, M.; Gabler, F.; Soding, J.; Lupas, A.N.; Alva, V. A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core. *J Mol Biol* **2018**, *430*, 2237-2243.