



## Genome Analysis of an Enterococcal Prophage, Entfac.MY

Maryam Yazdanizad<sup>1</sup>, Ramin Mazaheri Nezhad Fard<sup>2,3</sup>, Golshid Javdani Shahedin<sup>4</sup>,  
Mohammadreza Salehi<sup>5</sup>, Mahsa Dumanloo<sup>6</sup>, and Ali Akbar Saboor Yaraghi<sup>2\*</sup>

1. Department of Medical Biotechnology, Faculty of Advanced Science and Technology, Tehran Medical Sciences, Islamic Azad University, Tehran, Iran

2. Department of Pathobiology, School of Public Health, Tehran University of Medical Sciences, Tehran, Iran

3. Food Microbiology Research Center, Tehran University of Medical Sciences, Tehran, Iran

4. Pasteur Institute of Iran, Tehran, Iran

5. Department of Infectious Diseases, School of Medicine, Tehran University of Medical Sciences, Tehran, Iran

6. Department of Clinical Pathobiology, Imam Khomeini Hospital Complex, Tehran University of Medical Sciences, Tehran, Iran

### Abstract

**Background:** Bacteriophages are bacterial parasites. Unlike lytic bacteriophages, lysogenic bacteriophages do not multiply immediately after entering the host cells and may integrate their genomes into the bacterial genomes as prophages. Prophages can include various phenotypic and genotypic effects on the host bacteria. *Enterococcus* spp. are Gram-positive bacteria that cause infections in humans and animals. In recent decades, these bacteria have become resistant to various antimicrobials, including vancomycin. The aim of this study was to analyze genome of an enterococcal prophage.

**Methods:** In this study, *Enterococcus faecium* EntfacYE was isolated from biological samples and its genome was analyzed using next-generation sequencing method.

**Results:** Overall, 254 prophage genes were identified in the bacterial genome. The prophage included 39 housekeeping, 41 replication and regulation, 80 structural and packaging, and 48 lysis genes. Moreover, 46 genes with unknown functions were identified. All genes were annotated in DNA Data Bank of Japan.

**Conclusion:** In general, most prophage genes were linked to packaging and structure (31.5%) gene group. However, genes with unknown functions included a high proportion (18.11%), which indicated necessity of further analyses. Genomic analysis of the prophages can be effective in better understanding of their roles in development of bacterial resistance to antibiotics. Moreover, identification and study of prophages can help researchers develop genetic engineering tools and novel infection therapies.

*Avicenna J Med Biotech* 2022; 14(3): 196-205

**Keywords:** *Enterococcus faecium*, Genomic analysis, Prophage

\* **Corresponding author:**  
Ali Akbar Saboor Yaraghi, Ph.D.,  
Department of Pathobiology,  
School of Public Health, Tehran  
University of Medical Sciences,  
Tehran, Iran  
**Tel:** +98 21 42933168  
**Fax:** +98 21 88984913  
**E-mail:**  
asaboor@tums.ac.ir  
**Received:** 10 Nov 2021  
**Accepted:** 30 Apr 2022

### Introduction

Enterococci are Gram-positive bacteria found in soil, water, plants and dairy products. Furthermore, they are naturally found in regular gastrointestinal microbiota of the vertebrates. These bacteria may be detected on the skin and mucosa as well as in the mouth and vagina<sup>1,2</sup>. In addition, enterococci are commonly found in wastewater or fecal contaminated water<sup>3</sup>. Of the enterococcal species, *Enterococcus faecalis* (*E. faecalis*) and *Enterococcus faecium* (*E. faecium*) majorly infect humans<sup>4</sup>. Nowadays, antimicrobial resistance of *E. faecium* is a serious health threat according to the World Health Organization (WHO) and the

US Centers for Disease Control and Prevention (CDC). Enterococci increasingly show resistance to various antimicrobial drugs and disinfectants, including important antibiotic of vancomycin. This resistance prolongs the hospital stay and increases overall medical costs as well as risks of secondary infections and mortalities in patients<sup>2,5</sup>. Despite widespread phages in nature and their critical roles in medicine and industries, current knowledge of the biodiversity of these viruses is surprisingly small<sup>4</sup>.

Bacteriophages generally include two infection cycles of lytic and lysogenic cycles. In lytic cycle, phages

use host replication machinery to reproduce and eventually lyse the bacterial cells. In lysogenic cycle, phages integrate their genomes to the host genomes as prophages. Prophages may be conserved in the bacterial genomes for several years. Lysogenic phages enter the lytic phase (conversion) only when they are induced by natural factors (*e.g.* high temperatures) or laboratory agents (*e.g.* mitomycin C, hydrogen peroxide and UV radiation). In fact, conversion frequency of the prophages by natural inducers is low <sup>6</sup>. In general, study on the phages can help development of genetic engineering tools and novel infection therapies. Therefore, the major aim of the present study was to genetically analyze Entfac.MY prophage in a clinical isolate of *E. faecium*.

## Materials and Methods

### Bacterial strain

*E. faecium* was isolated from the biological sample of a hospitalized patient in Tehran, Iran, using routine microbiological methods. Bacterial isolation was verified using phenotypic methods such as Gram staining and biochemical methods such as catalase, oxidase, arabinose, bile esculin, NaCl, and PYR hydrolysis tests <sup>7</sup>. Additionally, the bacterial strain was genetically verified via amplification of the enterococcal *tuf* gene using PCR technique and specific primers of forward Ent1: 5'-TACTGACAAACCATTTCATGATG-3' and reverse Ent2: 5'-AACTTCGTCACCAACGCGAAC-3' <sup>8</sup>. The PCR cycling conditions were as follows: initial denaturation for 5 min at 95°C, followed by 30 cycles of denaturation at 95°C for 30 s, annealing at 54°C for 30 s and extension at 72°C for 45 s. Final extension was carried out at 72°C for 7 min (modified from Mazaheri Nezhad Fard *et al*, 2010). The PCR product was sequenced using Sanger method (Kowsar Biotech, Iran). Furthermore, antimicrobial susceptibility assessment of the isolate was carried out using disk diffusion method for vancomycin, erythromycin, clindamycin, ceftriaxone, and ceftioxin antibiotics. Characterized enterococcal strain was then used to isolate bacteriophages.

### Complete genome sequencing

In a previous study by Elahi *et al*, wastewater samples were used as the bacteriophage sources. After mixing 360 µl of each wastewater sample with 160 µl of the bacteria, suspension was poured onto a plate using two-layer agar method. Incubation was carried out at 37°C for 24 hr. After phage isolation on *E. faecium* strains, these bacterial strains were used for genome extraction <sup>9</sup>. A small quantity of agar (containing bacteria and phages) was dissolved in 1 ml of Saline Magnesium (SM) buffer. After 10 min of centrifugation at 4480 g, supernatant was separated and used for the bacterial genome extraction. Genome extraction was carried out using High Pure Viral Nucleic Acid Kit (Roche, Switzerland) as follows: 200 µl of working solution were added to a sterile microtube and mixed

with 50 µl of proteinase K. After addition of supernatant to the microtube, solution was mixed well using vortex and incubated at 72°C for 10 min. Then, 100 µl of binding buffer were added to the solution and mixed well. After transferring filter column into the microtube, 600 µl of the solution were added to the filter. Solution was centrifuged at 8000 g for 1 min at 4°C. Filter was transferred into a new microtube and 500 µl of inhibitor removal buffer were added to the filter. Solution was centrifuged at 8000 g for 1 min at 4°C. After transferring filter to a new microtube, 450 µl of wash buffer were added to the filter. Solution was centrifuged at 8000 g for 1 min at 4°C. Then, filter was transferred into a fresh microtube and the last two steps were repeated. To dry the filter, it was centrifuged at 13000 g for 10 s at 4°C. Filter was then transferred into a new sterile microtube and 30 µl of the elution buffer were added to the filter. After 1 min of setting at room temperature, centrifuge was carried out at 8000 g for 1 min at 4°C. The extracted bacterial genome was wholly sequenced using Illumina HiSeq Platform (Novogene, China) and analyzed using SPAdes *de novo* and reference assembly technologies. Technically, Illumina is a Next-Generation Sequencing (NGS) technology, which uses a proprietary reversible terminator-based technique that detects single bases as they are incorporated into the DNA template strands. In this study, prophage was analyzed using RAST online sequence analysis service and results were publicly deposited in DNA Data Bank of Japan (DDBJ).

## Results

Results of phenotypic and biochemical tests verified the primary characteristics of the enterococcal isolate. Gram-staining results showed that the isolate was a Gram-positive coccus. Results of the biochemical tests were reported as follows: negative catalase, negative oxidase, negative arabinose, positive bile esculin and positive PYR hydrolysis. Results of Sanger sequencing for the bacterial *tuf* gene also verified molecular characteristics of the enterococcal isolate (DDBJ accession numbers of LC580430 and LC580431). Antimicrobial susceptibility assessment results showed that *E. faecium* isolate was resistant to commonly used antibiotics, including vancomycin, erythromycin, clindamycin, ceftriaxone and ceftioxin. In general, Entfac.MY prophage included 130241 nucleotides with 35.74% A, 27.08% T, 16.02% C and 21.15% G. In total, 254 prophage genes were analyzed and grouped based on their functions (Table 1). These included 39 housekeeping genes, 41 replication and regulation genes, 80 packaging and structural genes, and 48 lysis group genes. Moreover, 46 genes with unknown functions were analyzed (Figure 1).

## Discussion

In general, bacteriophages isolated by exposing a clinical strain of *E. faecium* to wastewater samples

## Genome Analysis of Entfac.MY Prophage

Table 1. Genomic analysis of the Entfac.MY prophage and its similarity to other analyses from GenBank

CDS	Protein	Group	bp	DDBJ	GenBank	Cv	Id
1	Tape measure protein	PS	3849	LC644811	WP_002350712.1	99	100
2	YhgE/Pip domain-containing protein	H	2706	LC644652	WP_002296556.1	99	100
3	CHAP domain-containing protein	H	2316	LC644748	HAP5428988.1	98	99.21
4	DNA primase	RR	2313	LC644886	WP_187171400.1	99	99.87
5	Phage tail tape measure protein	PS	2310	LC644882	WP_002303051.1	99	100
6	Phage tail protein	H	1911	LC644883	BCR31617.1	100	99.84
7	Terminase large subunit	PS	1728	LC644732	WP_212463399.1	99	99.83
8	Phage major capsid protein	PS	1365	LC644734	MBJ1280756.1	99	99.78
9	Hypothetical protein	N/A	1359	LC644891	HAP4612701.1	100	99.78
10	Phage terminase large subunit	PS	1350	LC644750	WP_201689995.1	99	99.11
11	Terminase large subunit	PS	1320	LC644814	WP_010729419.1	99	100
12	Hypothetical protein	N/A	1296	LC644889	BCT02878.1	99	100
13	PcfJ family protein	N/A	1260	LC644747	WP_033782414.1	99	99.76
14	Phage portal protein	PS	1230	LC644881	AGS74854.1	99	99.76
15	Site-specific integrase	L	1227	LC644809	HAP7047036.1	99	99.75
16	Phage portal protein	PS	1221	LC644813	WP_002350718.1	99	100
17	Site-specific integrase	L	1215	LC644749	WP_002379954.1	99	100
18	Phage major capsid protein	PS	1194	LC644739	WP_219277563.1	99	99.75
19	Site-specific integrase	L	1182	LC644890	WP_002287705.1	96	100
20	Site-specific integrase	L	1179	LC644950	WP_002288969.1	99	100
21	Site-specific integrase	L	1143	LC631094	MBK8154746.1	73.56	0.0
22	Site-specific integrase	L	1140	LC644714	WP_073470529.1	99	99.74
23	Site-specific integrase	L	1137	LC644817	WP_012197643.1	96	100
24	Site-specific integrase	L	1128	LC644951	WP_002321493.1	98	100
25	Site-specific integrase	L	1122	LC644744	HBB6776700.1	98	100
26	Portal protein	H	1035	LC631113	WP_105116343.1	97	52.66
27	DNA polymerase	RR	1022	LC644897	HAP4621056.1	99	99.71
28	N-acetylmuramoyl-L-alanine amidase	L	1017	LC644737	WP_020944989.1	96	99.69
29	DNA polymerase	RR	1008	LC644820	NSU25233.1	92	100
30	Phage tail protein	H	972	LC644761	WP_165548432.1	100	100
31	YqaJ viral recombinase family protein	RR	942	LC644745	WP_033646797.1	99	100
32	DUF2800 domain-containing protein	N/A	924	LC644831	HAP3631774.1	76	99.15
33	Phage baseplate upper protein	PS	906	LC644824	WP_048946725.1	99	100
34	Deoxyribonuclease IV	RR	903	LC644819	MBR8696505.1	99	99.67
35	Hypothetical protein	N/A	894	LC644752	HAP5478159.1	98	96.99
36	Recombinase RecT	RR	891	LC644746	WP_002301573.1	99	100
37	Hypothetical protein	N/A	840	LC644899	HB11605865.1	100	99.64
38	DEAD/DEAH box helicase family protein	RR	837	LC644756	WP_185933038.1	100	100
39	Small subunit of terminase	PS	834	LC644888	WP_048946579.1	99	100
40	N-acetylmuramoyl-L-alanine amidase	L	825	LC644836	HBA0027879.1	100	100
41	LysM peptidoglycan-binding domain-containing protein	H	813	LC644828	HAP5669678.1	94	100
42	YhgE/Pip domain-containing protein	H	810	LC644765	WP_195947705.1	91	100
43	Hypothetical protein	N/A	789	LC644770	EHU5028596.1	77	95.51
44	Phage antirepressor	RR	777	LC644806	WP_002290310.1	99	100
45	Helix-turn-helix domain-containing protein	RR	750	LC644730	WP_130017038.1	99	100
46	Phage major capsid protein	PS	738	LC644900	WP_153883634.1	100	100
47	Phage portal protein	PS	732	LC644829	WP_195960918.1	99	97.94
48	YhgE/Pip domain-containing protein	H	726	LC644841	WP_074934422.1	90	100
49	Phage regulatory protein	RR	723	LC644953	WP_002403884.1	99	98.75
50	Phage tail tape measure protein	PS	723	LC644832	HAP3877100.1	100	99.59
51	DUF4145 domain-containing protein	N/A	714	LC631111	WP_033585566.1	96	57.33
52	ORF6C domain-containing protein	H	698	LC644715	WP_213386100.1	98	97.25
53	Phage tail tape measure protein	PS	696	LC644843	WP_010773843.1	100	100
54	Deoxyribonuclease IV	RR	693	LC644912	MBS6508197.1	100	100
55	AAA family ATPase	H	687	LC631112	WP_000704947.1	95	52.47
56	Phage tail protein	H	681	LC644913	WP_198765169.1	100	98.68
57	YhgE/Pip domain-containing protein	H	666	LC644769	EOI23661.1	84	97.34
58	Antireceptor	H	660	LC631105	APC45840.1	59	47.47
59	Hypothetical protein	N/A	657	LC644755	MBS5814550.1	99	100
60	YhgE/Pip domain-containing protein	H	657	LC644846	WP_053832241.1	100	100
61	Hypothetical protein	N/A	648	LC631102	WP_044981200.1	82	64.25
62	Peptidoglycan recognition protein	RR	639	LC644844	MBJ1172703.1	100	100
63	Phage tail protein	H	627	LC644736	MBG7868557.1	99	99.04
64	Site-specific integrase	L	627	LC644847	WP_137663438.1	100	100
65	YhgE/Pip domain-containing protein	H	624	LC644840	RDC50037.1	100	100
66	Tail tape measure protein	PS	615	LC644916	EGO8897385.1	100	99.02
67	Lysin	L	609	LC644835	EGO8925797.1	99	100
68	YhgE/Pip domain-containing protein	H	606	LC644848	WP_192380427.1	100	99.01
69	Tape measure protein	PS	600	LC644918	HAP2857181.1	100	100
70	Recombinase family protein	RR	594	LC644774	EHV0135166.1	100	100

Contd. Table 1. Genomic analysis of the Enfac.MY prophage and its similarity to other analyses from GenBank

CDS	Protein	Group	bp	DDBJ	GenBank	Cv	Id
71	Tape measure protein	PS	591	LC644902	HBC4147789.1	98	97.95
72	YhgE/Pip domain-containing protein	H	591	LC644909	WP_142424143.1	100	100
73	Phage tail tape measure protein	PS	585	LC644778	WP_010773806.1	100	99.49
74	Thermonuclease family protein	RR	585	LC644821	WP_002368095.1	99	100
75	Structural protein	PS	585	LC644919	HAP4242704.1	97	97.89
76	Recombinase family protein	RR	582	LC644651	WP_002294320.1	99	100
77	DUF2815 family protein	N/A	576	LC631101	WP_199363988.1	88	60.95
78	HK97 family phage prohead protease	PS	570	LC644733	WP_002303035.1	99	100
79	Endonuclease	L	570	LC644911	EGO6070115.1	99	100
80	Deoxyribonuclease IV	RR	570	LC644920	WP_048952159.1	100	100
81	Phage tail tape measure protein	PS	564	LC644921	HAP3877078.1	100	100
82	Phage tail protein	H	561	LC644812	WP_113998124.1	98	99.46
83	Phage scaffolding protein	PS	558	LC631103	WP_171943061.1	77	41.67
84	Tyrosine-type recombinase/integrase	L	555	LC644818	WP_204357435.1	99	100
85	DEAD/DEAH box helicase	RR	534	LC644898	HAP5399412.1	99	100
86	Hypothetical protein	N/A	525	LC631114	WP_123159857.1	92	69.57
87	Phage major capsid protein	PS	516	LC644772	EGO2693906.1	100	100
88	Phage portal protein	PS	513	LC644833	MRI70555.1	100	95.91
89	HK97 family phage prohead protease	PS	510	LC644775	WP_086295712.1	99	98.82
90	Terminase large subunit	PS	495	LC644924	HBI1611120.1	100	100
91	Recombinase RecT	RR	480	LC644763	WP_142974208.1	99	100
92	Terminase large subunit	PS	474	LC644741	WP_154059379.1	93	100
93	Tyrosine-type recombinase/integrase	L	474	LC631100	WP_153046652.1	60	56.84
94	Terminase large subunit	PS	474	LC644915	NSP35552.1	99	100
95	Recombinase family protein	RR	471	LC644766	WP_035006379.1	99	99.36
96	Helix-turn-helix transcriptional regulator	RR	471	LC644839	HAP4862478.1	99	100
97	Hypothetical protein	N/A	459	LC644764	WP_077143720.1	100	100
98	Site-specific integrase	L	459	LC644925	WP_194193259.1	100	100
99	Tyrosine-type recombinase/integrase	L	456	LC644781	WP_086295839.1	98	100
100	P27 family phage terminase small subunit	PS	453	LC644731	HAQ5377436.1	99	94.67
101	Phage tail family protein	H	453	LC644782	WP_142968761.1	100	100
102	Hypothetical protein	N/A	450	LC644751	WP_002298034.1	99	100
103	Peptidoglycan endopeptidase EnpA	L	441	LC644856	WP_010706628.1	100	100
104	Phage major capsid protein	PS	438	LC644857	WP_216442317.1	100	99.32
105	Hypothetical protein	N/A	435	LC644849	WP_002386667.1	100	100
106	Site-specific integrase	L	429	LC644773	EGO5860053.1	100	97.90
107	Phage tail tape measure protein	PS	429	LC644784	EHV2895864.1	100	100
108	Siphovirus Gp157 family protein	N/A	420	LC631099	WP_048784131.1	99	50.00
109	Autolysin	L	414	LC644815	GER95770.1	99	100
110	Hypothetical protein	N/A	414	LC644930	WP_195960930.1	86	98.33
111	Phage major capsid protein	PS	405	LC631108	WP_096649614.1	97	55.30
112	Site-specific integrase	L	402	LC644859	WP_165710100.1	100	98.51
113	Phage tail protein	H	396	LC644787	EHR4559110.1	100	98.48
114	Hypothetical protein	N/A	396	LC644838	WP_048943143.1	99	99.24
115	Hypothetical protein	N/A	396	LC644885	WP_002350714.1	99	100
116	Phage tail tape measure protein	PS	390	LC644861	EHV2895864.1	100	96.92
117	Recombinase RecT	RR	387	LC644780	WP_002380492.1	99	100
118	Hypothetical protein	N/A	387	LC631109	WP_102570448.1	40	50.00
119	RloB domain-containing protein	H	387	LC631116	WP_003011523.1	88	37.72
120	Baseplate J/gp47 family protein	PS	381	LC644788	EGS7987863.1	100	100
121	HNH endonuclease	L	381	LC644880	HBC4262174.1	99	100
122	Tail tape measure protein	PS	378	LC644789	HAP4230866.1	100	100
123	Helix-turn-helix transcriptional regulator	RR	375	LC644743	HAQ4376708.1	99	99.19
124	Phage tail tape measure protein	PS	375	LC644757	EGO8777387.1	99	99.19
125	Hypothetical protein	N/A	375	LC644952	WP_048943142.1	99	100
126	Toxin	N/A	375	LC644826	NSU57868.1	99	85.48
127	HNH endonuclease	L	366	LC644767	WP_086269578.1	99	99.17
128	Phage tail tape measure protein	PS	363	LC644790	MRI75296.1	100	99.17
129	Deoxyribonuclease IV	RR	363	LC644791	WP_201688942.1	100	100
130	YhgE/Pip domain-containing protein	H	363	LC644792	WP_167822678.1	100	97.52
131	DUF1073 domain-containing protein	N/A	363	LC644863	WP_010706623.1	100	100
132	Terminase large subunit	PS	363	LC644908	WP_086295714.1	99	100
133	Phage major capsid protein	PS	360	LC644933	WP_216442317.1	100	100
134	Terminase large subunit	PS	357	LC644932	WP_086295714.1	100	97.48
135	Phage tail family protein	H	351	LC644934	WP_048948486.1	100	100
136	PBSX family phage terminase large subunit	PS	345	LC644868	HBD0933441.1	100	100
137	Phage tail protein	H	342	LC644768	HBD0803692.1	100	100
138	Siphovirus ReqiPepy6 Gp37-like family protein	N/A	342	LC644928	WP_048943146.1	100	100
139	Hypothetical protein	N/A	339	LC644808	WP_010706753.1	100	100
140	Hypothetical protein	N/A	339	LC644822	WP_002303043.1	99	100



## Genome Analysis of Entfac.MY Prophage

Contd. Table 1. Genomic analysis of the Entfac.MY prophage and its similarity to other analyses from GenBank

CDS	Protein	Group	bp	DDBJ	GenBank	Cv	Id
141	Nucleoid-associated protein	RR	330	LC644870	WP_021428623.1	87	100
142	Toxin	N/A	321	LC644910	HAP4862477.1	100	100
143	Site-specific integrase	L	321	LC631118	MBK8154746.1	97	91.35
144	VRR-NUC domain-containing protein	H	318	LC644816	WP_002350665.1	99	99.05
145	Lysozyme family protein	L	318	LC644937	WP_156233373.1	93	100
146	MazG-like family protein	RR	315	LC644742	WP_012197627.1	99	100
147	Helix-turn-helix domain-containing protein	RR	312	LC631107	WP_118138944.1	99	52.38
148	DNA methyltransferase	RR	309	LC644794	WP_195960921.1	100	99.03
149	Hypothetical protein	N/A	309	LC644866	WP_002369984.1	100	100
150	Phage portal protein	PS	309	LC644896	WP_201689987.1	100	99.03
151	Phage tail tape measure protein	PS	306	LC644871	WP_153841946.1	100	98.04
152	Hypothetical protein	N/A	303	LC644807	HAQ9461605.1	99	99
153	Phage tail tape measure protein	PS	300	LC644867	EGO8777387.1	80	100
154	Hypothetical protein	N/A	300	LC644895	EGO5085241.1	99	96.97
155	Hypothetical protein	N/A	300	LC644938	WP_010748256.1	85	100
156	Tyrosine-type recombinase/integrase	L	297	LC644939	WP_216442824.1	100	100
157	Terminase large subunit	PS	297	LC631119	MBF1717610.1	100	77.78
158	Hypothetical protein	N/A	294	LC644907	WP_096709598.1	98	78.35
159	HK97 family phage prohead protease	PS	293	LC644837	WP_137192767.1	99	100
160	Nucleoid-associated protein	RR	291	LC644777	WP_035114676.1	100	100
161	Site-specific integrase	L	291	LC644796	EGO9032817.1	100	100
162	Peptidoglycan endopeptidase EnpA	L	291	LC644922	WP_002400013.1	88	100
163	Site-specific integrase	L	285	LC631096	WP_128080223.1	86	71.95
164	AAA family ATPase	H	285	LC644873	WP_010828176.1	98	100
165	Phage gp6-like head-tail connector protein	PS	282	LC644735	WP_002303039.1	98	100
166	YhgE/Pip domain-containing protein	H	282	LC644834	WP_142966911.1	98	100
167	Phage gp6-like head-tail connector protein	PS	282	LC644851	WP_161843549.1	98	95.70
168	Tyrosine-type recombinase/integrase	L	282	LC644865	WP_201705518.1	100	97.87
169	Phage terminase large subunit	PS	282	LC644940	EGO6069701.1	100	100
170	Site-specific integrase	L	279	LC644771	WP_192201846.1	100	100
171	Terminase large subunit	PS	276	LC644799	EHU8855151.1	100	100
172	Minor capsid protein	PS	276	LC644850	WP_142425073.1	98	100
173	Phage portal protein	PS	273	LC631097	APD22584.1	100	54.95
174	Structural protein	PS	267	LC644759	EGO2667974.1	98	100
175	Phage tail tape measure protein	PS	264	LC644800	HAP4877189.1	90	98.75
176	YhgE/Pip domain-containing protein	H	264	LC644923	HAP5737370.1	100	100
177	Phage tail tape measure protein	PS	261	LC644874	EHB6454364.1	100	89.66
178	Phage terminase large subunit	PS	261	LC644936	HAP5353043.1	98	97.67
179	Phage portal protein	PS	258	LC644802	WP_194943338.1	100	100
180	Phage major capsid protein	PS	258	LC644935	WP_002417392.1	98	100
181	Hemolysin XhIA family protein	L	255	LC644758	WP_048948902.1	98	92.86
182	Hypothetical protein	N/A	255	LC631098	WP_018165267.1	100	44.94
183	MazG-like family protein	RR	255	LC644901	WP_010717138.1	100	100
184	Site-specific integrase	L	252	LC644853	HAP4918633.1	100	100
185	Hypothetical protein	N/A	249	LC644893	WP_010748256.1	100	96.39
186	Phage tail protein	H	249	LC644825	MRJ04801.1	91	82.89
187	Hypothetical protein	N/A	249	LC644803	HAP2781243.1	98	100
188	AAA family ATPase	H	246	LC644943	WP_127341875.1	75	100
189	Phage tail tape measure protein	PS	246	LC644944	WP_118215595.1	100	95.12
190	Hemolysin XhIA family protein	L	243	LC644810	WP_002332427.1	98	100
191	Phage tail tape measure protein	PS	243	LC644876	WP_085406750.1	100	100
192	Hypothetical protein	N/A	243	LC644929	MRJ04805.1	98	100
193	Helix-turn-helix transcriptional regulator	RR	240	LC644779	HAP3299532.1	100	100
194	Phage tail tape measure protein	PS	237	LC644947	MBR8697128.1	68	100
195	Phage tail tape measure protein	PS	237	LC644804	MRI62301.1	100	98.73
196	Hemolysin XhIA family protein	L	234	LC644753	WP_214144888.1	98	96.10
197	Collagen-like protein	PS	234	LC644776	WP_207656718.1	98	97.40
198	Phage tail protein	H	234	LC644786	EHB6443012.1	98	100
199	Peptidoglycan endopeptidase EnpA	L	234	LC644872	WP_138218550.1	100	100
200	Phage terminase large subunit	PS	234	LC644877	HBD0837071.1	100	100
201	Terminase large subunit	PS	234	LC631117	MBF0843203.1	98	59.74
202	Hypothetical protein	N/A	231	LC631115	WP_125852315.1	100	41.56
203	Terminase small subunit	PS	231	LC644926	WP_070363956.1	98	100
204	Site-specific integrase	L	228	LC631110	WP_044761187.1	98	60.00
205	DUF2213 domain-containing protein	N/A	228	LC644879	WP_115252339.1	100	100
206	Site-specific integrase	L	228	LC644949	EGT2099792.1	100	100
207	YhgE/Pip domain-containing protein	H	225	LC644805	RDC49008.1	100	100
208	Phage holin	RR	225	LC631106	WP_155198660.1	69	98.08
209	DUF3383 family protein	N/A	225	LC644878	WP_002377529.1	100	100
210	IS607 family transposase	RR	225	LC644887	WP_104845943.1	90	54.17

Contd. Table 1. Genomic analysis of the Enfac.MY prophage and its similarity to other analyses from GenBank

CDS	Protein	Group	bp	DDBJ	GenBank	Cv	Id
211	Phage tail protein	H	225	LC644931	WP_198765584.1	100	98.67
212	Terminase large subunit	PS	222	LC644855	HBII611155.1	100	100
213	Phage portal protein	PS	216	LC644842	EOI11687.1	100	98.61
214	Phage holin	RR	216	LC644903	WP_033626583.1	98	100
215	Holin	RR	207	LC644754	WP_002369080.1	97	100
216	Phage major tail protein	PS	207	LC644858	EFT89608.1	100	100
217	Baseplate J/gp47 family protein	PS	207	LC644864	HAP3865876.1	100	98.55
218	Phage holin	RR	198	LC644738	WP_002333030.1	98	98.46
219	Cupin domain-containing protein	H	192	LC644845	WP_168932514.1	98	98.41
220	HNH endonuclease	L	192	LC644927	EHU8855159.1	100	100
221	AAA family ATPase	H	189	LC644860	WP_127341875.1	100	98.41
222	Phage major capsid protein	PS	186	LC644795	NSR35780.1	95	96.61
223	Hypothetical protein	N/A	186	LC644894	APS16315.1	98	100
224	1,4-β-N-acetylmuramidase	L	186	LC644942	HAP4707744.1	100	100
225	Ig domain-containing protein	H	183	LC644905	WP_010773805.1	98	95.00
226	Structural protein	PS	174	LC644760	MBO6404911.1	100	98.28
227	Hypothetical protein	N/A	168	LC631104	WP_050285512.1	83	53.19
228	Site-specific integrase	L	165	LC644948	EGO8320093.1	98	100
229	Phage holin	RR	162	LC644827	WP_002379919.1	100	100
230	Ribbon-helix-helix domain-containing protein	RR	159	LC644740	WP_010729417.1	98	100
231	DUF2800 domain-containing protein	N/A	159	LC644892	WP_198765184.1	98	100
232	Deoxyribonuclease IV	RR	153	LC644801	WP_144330580.1	98	100
233	Phage tail tape measure protein	PS	147	LC644906	HAP2938854.1	100	100
234	Terminase large subunit	PS	144	LC644946	HBII611197.1	97	100
235	DUF4065 domain-containing protein	N/A	141	LC644793	HAP4215483.1	97	100
236	XkdX family protein	L	138	LC644862	WP_071974452.1	97	100
237	XkdX family protein	L	138	LC644884	WP_002332428.1	97	100
238	Phage tail tape measure protein	PS	138	LC644904	MBO6395993.1	97	100
239	N-acetylmuramoyl-L-alanine amidase	L	135	LC644798	WP_192400940.1	91	90.24
240	Nucleoid-associated protein	RR	132	LC644783	WP_114543333.1	95	95.24
241	Hypothetical protein	N/A	126	LC644762	WP_086295717.1	100	97.62
242	Ig domain-containing protein	H	120	LC644785	WP_048943144.1	100	100
243	HNH endonuclease	L	114	LC644914	NSR35697.1	97	100
244	DUF2184 domain-containing protein	N/A	108	LC644854	WP_002386753.1	100	100
245	Phage major capsid protein	PS	102	LC644917	WP_086295711.1	100	100
246	DUF2184 domain-containing protein	N/A	96	LC644797	WP_010706626.1	96	100
247	Tyrosine-type recombinase/integrase	L	84	LC644852	WP_153841893.1	96	100
248	Helix-turn-helix domain-containing protein	RR	81	LC631095	WP_048782931.1	100	85.19
249	BppU family phage baseplate upper protein	PS	78	LC644830	HAP6528242.1	100	100
250	Phage head-tail connector protein	H	78	LC644869	WP_214144901.1	100	100
251	Phage portal protein	PS	78	LC644945	WP_114679410.1	100	96.15
252	XkdX family protein	L	75	LC644823	WP_074394623.1	100	100
253	Phage major capsid protein	PS	75	LC644941	HAP2938884.1	100	100
254	Holin	RR	69	LC644875	HBC4422148.1	95	100

CDS: Coding Sequence, *bp*: Base Pair, DDBJ: DNA Data Bank of Japan, GenBank: GenBank Protein ID for this study, Cv: Coverage, Id: Identity, N/A: Not Applicable or Announced, L: Lysis, PS: Packaging and structural, H: Housekeeping, RR: Replication and regulation.

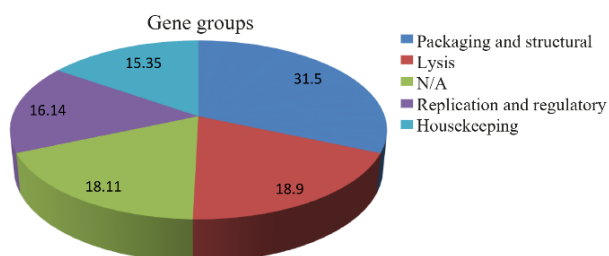


Figure 1. Major gene groups of Enfac.MY prophage.

included lysogenic phages. After complete genomic sequencing of Entfac.MY prophage, 254 genes were functionally analyzed. The housekeeping group of

genes included YhgE/Pip domain-containing protein (12 copies), CHAP domain-containing protein, phage tail protein (ten copies), RloB domain-containing protein, antireceptor, LysM peptidoglycan-binding domain-containing protein, ORF6C domain-containing protein, VRR-NUC domain-containing protein, Ig domain-containing protein (two copies), portal protein (ten copies), AAA family ATPase (four copies), phage tail family protein (two copies) and phage head-tail connector protein. Portal protein forms a channel for two-way passes of the viral DNA. The EFRM31 phage genome also included this gene <sup>10</sup>. The functional mechanisms of AAA family ATPase, including cell cycle regulation, proteolysis and protein breakdown and intracellular transport, were investigated in 2002 <sup>11</sup>.

The phage tail family protein encodes tail components. The antireceptor gene is involved in identifying and binding phage to its bacterial hosts. Function of this gene was previously reported by Vegge *et al*<sup>12</sup> and Duplessis and Moineau<sup>13</sup>. The phage head-tail connector protein gene is involved in the assembly of virions by connecting their heads and tails.

The replication and regulation group of genes included IS607 family transposase, ribbon-helix-helix domain-containing protein, recombinase RecT (three copies), phage regulatory protein, thermonuclease family protein, helix-turn-helix transcriptional regulator (three copies), MazG-like family protein (two copies), DNA polymerase (two copies), DNA primase, deoxyribonuclease IV (five copies), recombinase family protein (four copies), DEAD/DEAH box helicase family protein, phage antirepressor, helix-turn-helix domain-containing protein (three copies), peptidoglycan recognition protein, nucleoid-associated protein (three copies), DNA methyltransferase and phage holin (six copies). In fact, DNA polymerase gene produces enzymes that act in pairs to produce two identical DNA strands from an original DNA molecule and are essential for DNA replication. Krzywkowski *et al* studied this gene in 2018<sup>14</sup>. Naturally, DNA primase is an enzyme involved in DNA replication and is linked to RNA polymerases. Structure and mechanism of this gene have been studied by O'Brien *et al* in 2018<sup>15</sup> and Lee *et al* in 2010<sup>16</sup>. When cells are in repairing process of the nucleotide cleavage pathway, deoxyribonuclease IV is active in aqueous or aporin-apyrimidinic sites. The gene function has been explained by Enriquez *et al*<sup>17</sup>. Recombinase family protein produces enzymes involved in repairing DNA damages. Voziyanova *et al* investigated function and evolution of this gene<sup>18</sup>. The DEAD/DEAD box helicase family protein produces enzymes that bind and hydrolyze NTP to double-stranded (ds) RNA molecules or regenerate RNA protein complexes. The phage antirepressor gene prevents suppressor proteins from binding to their operators. This mechanism has recently been investigated by Silpe *et al* in 2020<sup>19</sup>. The helix-turn-helix domain-containing protein can bind to DNA molecules. The peptidoglycan recognition protein gene detects peptidoglycan in the bacterial cell walls; as described by Skerry *et al*<sup>20</sup> and Jiang *et al*<sup>21</sup>. Nucleoid-associated proteins naturally include abundant polypeptides with Low-Molecular Weights (LMW) that bind DNA molecules and change their shapes and abilities to participate in processes such as transcription. Sometimes, these proteins can bind RNA molecules and post-transcriptionally affect the gene expression of cells; as described by Stojkova *et al* in 2019<sup>22</sup>. DNA methyltransferases are a family of enzymes involved in production and maintenance of CpG methylation in the genome; as described by Furuta *et al* in 2014<sup>23</sup> and Stoddard *et al* in 2019<sup>24</sup>. Holins include diverse groups of small proteins from dsDNA phages that destruct the

host cell walls in the virus lytic cycle. Functional mechanisms of the holins were previously described by Bardina *et al*<sup>25</sup>, Stamereilers *et al*<sup>26</sup> and Jacobs *et al*<sup>27</sup>.

The packaging and structure group of genes included tail tape measure protein (24 copies), P27 family phage terminase small subunit, baseplate J/gp47 family protein (two copies), PBSX family phage, phage gp6-like head-tail connector protein (two copies), collagen-like protein, phage major tail protein, BppU family phage baseplate upper protein, HK97 family phage prohead protease (three copies), phage scaffolding protein, terminase large subunit (17 copies), phage major capsid protein (11 copies), minor capsid protein, phage baseplate upper protein, structural protein (three copies) and small subunit of terminase (two copies). The terminase large subunit gene is involved in viral DNA transfer and packaging termination. The phage major capsid protein gene encodes capsid proteins and was previously reported in the phage genome of EFRM31 in 2010<sup>10</sup>. Stamereilers *et al* reported this gene as well<sup>26</sup>. Another gene involved in encoding capsid proteins is the minor capsid protein gene. In 2015, Pawlowski *et al* studied functions of this gene in P23-77 phage<sup>28</sup>. The phage baseplate upper protein gene was also reported by Li *et al* in  $\phi$ 11 genome<sup>29</sup>. Structural protein gene plays an important role in shaping structure of the virus. In 2015, McNulty *et al* studied terminase small subunit gene, which is responsible for binding of the encoded protein to several identifying elements at the beginning of viral packaging<sup>30</sup>. In addition, Roy *et al* studied this gene in 2011<sup>31</sup>.

The lysis group of genes included hemolysin Xh1A family protein (three copies), 1,4- $\beta$ -N-acetylmuramidase, XkdX family protein (three copies), site-specific integrase (21 copies), N-acetylmuramoyl-L-alanine amidase (three copies), lysin, endonuclease, tyrosine-type recombinase/integrase (six copies), peptidoglycan endopeptidase EnpA (three copies), autolysin, HNH endonuclease (four copies) and lysozyme family protein. Site-specific integrase gene is responsible for rearranging DNA fragments; as previously identified by Petersen *et al* in the genome of TPW22 phage<sup>32</sup>. This gene was also identified in TP901-1 phage genome<sup>33</sup>. The N-acetylmuramoyl-L-alanine amidase gene produces an enzyme that catalyzes a chemical reaction and cleaves the link between N-acetylmuramoyl and L-amino acid residues in the cell-wall glycopeptides. Bierbaum *et al* have described functional mechanisms of this gene<sup>34</sup>. Lysines are hydrolytic phage enzymes produced to separate the host cell walls in lytic cycles. In 2015, LeBlanc *et al* reported this gene in their phage genome<sup>35</sup>. The endonuclease gene produces enzymes that break down DNA molecules at specific locations. The tyrosine-type recombinase/integrase gene is involved in DNA binding and recombination. In a similar study in 2021, Malecki *et al* reported peptidoglycan endopeptidase EnpA gene in an enterococcal prophage

<sup>36</sup>. The autolysin gene is an enzyme that breaks down the components of peptidoglycans in cells and separates daughter cells after cell division. Ju *et al* in 2012 <sup>37</sup> and Kohler *et al* in 2014 studied this gene <sup>38</sup>. The HNH endonuclease gene plays various roles in the phage life cycle as the major component of the phage DNA packaging machinery. In 2017, Zhang *et al* studied the mechanism of this gene in GVE2 phage genome <sup>39</sup>. Lysozyme is a glycoside hydrolase that catalyzes the hydrolysis of 1,4- $\beta$  bonds between N-acetylmuramic and N-acetyl-D-glucosamine residues of the peptidoglycan. Previously, Irwin studied this gene in 2014 <sup>40</sup>.

The unknown function group of genes included PcfJ family protein, DUF2800 domain-containing protein, DUF4145 domain-containing protein, DUF2815 family protein, siphovirus Gp157 family protein, toxin (two copies), DUF1073 domain-containing protein, siphovirus ReqiPepy6 Gp37-like family protein, DUF2213 domain-containing protein, DUF3383 family protein, DUF4065 domain-containing protein, DUF2184 domain-containing protein (two copies) and hypothetical protein (31 copies). Hypothetical proteins were previously addressed by Mazaheri *et al* in EFRM31 <sup>10</sup> and Tang *et al* in  $\phi$ NJ2 <sup>41</sup>. Based on the complete genomic analysis of an enterococcal phage by Mazaheri Nezhad Fard *et al* in 2010, genes similar to the genes of this study were reported, including hypothetical proteins, portal proteins, major capsid proteins, and holins <sup>10</sup>. In another study by Tan *et al* in 2007, genes similar to the genes of the current study were reported as well, including terminase large subunit and portal protein genes <sup>42</sup>. In a study by O'Flaherty *et al* on staphylococcal phage genomes (2004), 63 hypothetical protein genes were identified <sup>43</sup> while 31 hypothetical protein genes were identified in the present study. In the present study, AAA family ATPase, holin, and major capsid protein genes were successfully characterized.

### Conclusion

In general, bacteriophages are present in all environments with alive bacteria, especially in wastewaters. Bacteriophages can insert their genomes into the bacterial genomes to become prophages. Prophages can regulate bacterial populations by changing the bacterial gene expression rates and are involved in bacterial resistance by transferring antibiotic resistance genes to their bacterial hosts. Up-to-date, pathogenic bacteria have developed multiple resistance to available antibiotics, creating serious problems with costly infection treatments and unwanted mortalities. Bacteriophages can be used as novel solutions for these problems. Therefore, genomic analysis of the prophages can improve better understanding of their effects on bacteria. Furthermore, study of prophages can help researchers develop novel genetic engineering tools and effective medical therapies.

### Acknowledgement

The authors thank staff within the Microbiology Laboratory, Department of Pathobiology, School of Public Health, Tehran University of Medical Sciences.

### Conflict of Interest

The authors declare no conflict of interest.

### Ethics Statement

The current study was carried out based on the methods approved by the Ethics Committee of Tehran University of Medical Sciences (IR.TUMS.SPH.REC.1397.139).

### References

1. Fisher K, Phillips C. The ecology, epidemiology and virulence of *Enterococcus*. *Microbiology* 2009;155(Pt 6):1749-57.
2. Garcia-Solache M, Rice LB. The *Enterococcus*: a model of adaptability to its environment *Clin Microbiol Rev* 2019;32(2):e00058-18.
3. Zaheer R, Cook SR, Barbieri R, Goji N, Cameron A, Petkau A, et al. Surveillance of *Enterococcus* spp. reveals distinct species and antimicrobial resistance diversity across a One-Health continuum. *Sci Rep* 2020;10(1):3937.
4. Topka-Bielecka G, Bloch S, Nejman-Falenczyk B, Grabski M, Jurczak-Kurek A, Gorniak M, et al. Characterization of the bacteriophage vB\_EfaS-271 infecting *Enterococcus faecalis*. *Int J Mol Sci* 2020;21(17):6345.
5. Revtovich AV, Tjahjono E, Singh KV, Hanson BM, Murray BE, Kirienko NV. Development and characterization of high-throughput *Caenorhabditis elegans*-*Enterococcus faecium* infection model. *Front Cell Infect Microbiol* 2021;11:667327.
6. Abdulrahman Ashy R, Agusti S. Low host abundance and high temperature determine switching from lytic to lysogenic cycles in planktonic microbial communities in a tropical sea (Red Sea). *Viruses* 2020;12(7):761.
7. Stepien-Pysniak D, Hauschild T, Dec M, Marek A, Urban-Chmiel R, Kosikowska U. Phenotypic and genotypic characterization of *Enterococcus* spp. from yolk sac infections in broiler chicks with a focus on virulence factors. *Poult Sci* 2021;100(4):100985.
8. Nasiri M, Hanifian S. *Enterococcus faecalis* and *Enterococcus faecium* in pasteurized milk: Prevalence, genotyping, and characterization of virulence traits. *LWT* 2022; 153:112452.
9. Elahi Y, Nowroozi J, Fard RMN. Isolation and characterization of bacteriophages from wastewater sources on *Enterococcus* spp. isolated from clinical samples. *Iran J Microbiol* 2021;13(5):671-7.
10. Fard RM, Barton MD, Arthur JL, Heuzenroeder MW. Whole-genome sequencing and gene mapping of a newly isolated lytic enterococcal bacteriophage EFRM31. *Arch Virol* 2010;155(11):1887-91.
11. Elderkin S, Jones S, Schumacher J, Studholme D, Buck



- M. Mechanism of action of the Escherichia coli phage shock protein PspA in repression of the AAA family transcription factor PspF. *J Mol Biol* 2002;320(1):23-37.
12. Vegge CS, Vogensen FK, Mc Grath S, Neve H, van Sinderen D, Brøndsted L. Identification of the lower baseplate protein as the antireceptor of the temperate lactococcal bacteriophages TP901-1 and Tuc2009. *J Bacteriol* 2006;188(1):55-63.
  13. Duplessis M, Moineau S. Identification of a genetic determinant responsible for host specificity in Streptococcus thermophilus bacteriophages. *Mol Microbiol* 2001;41(2):325-36.
  14. Krzywkowski T, Kuhnemund M, Wu D, Nilsson M. Limited reverse transcriptase activity of phi29 DNA polymerase. *Nucleic Acids Res* 2018;46(7):3625-32.
  15. O'Brien E, Holt ME, Salay LE, Chazin WJ, Barton JK. Substrate binding regulates redox signaling in human DNA primase. *J Am Chem Soc* 2018;140(49):17153-62.
  16. Lee SJ, Zhu B, Hamdan SM, Richardson CC. Mechanism of sequence-specific template binding by the DNA primase of bacteriophage T7. *Nucleic Acids Res* 2010;38(13):4372-83.
  17. Enriquez A, Chu IW, Mellis C, Lin WY. Nebulised deoxyribonuclease for viral bronchiolitis in children younger than 24 months. *Cochrane Database Syst Rev* 2012;(11): 11(11):CD008395.
  18. Voziyanova E, Li F, Shah R, Voziyanov Y. Genome targeting by hybrid FIp-TAL recombinases. *Sci Rep* 2020;10(1):17479.
  19. Silpe JE, Bridges AA, Huang X, Coronado DR, Duddy OP, Bassler BL. Separating functions of the phage-encoded quorum-sensing-activated antirepressor Qtip. *Cell Host Microbe* 2020;27(4):629-41.
  20. Skerry C, Goldman WE, Carbonetti NH. Peptidoglycan recognition protein 4 suppresses early inflammatory responses to Bordetella pertussis and contributes to sphingosine-1-phosphate receptor agonist-mediated disease attenuation. *Infect Immun* 2019;87(2):e00601-18.
  21. Jiang L, Liu W, Guo H, Dang Y, Cheng T, Yang W, et al. Distinct functions of Bombyx mori peptidoglycan recognition protein 2 in immune responses to bacteria and viruses. *Front Immunol* 2019;10:776.
  22. Stojkova P, Spidlova P, Stulik J. Nucleoid-associated protein HU: A lilliputian in gene regulation of bacterial virulence. *Front Cell Infect Microbiol* 2019;9:159.
  23. Furuta Y, Namba-Fukuyo H, Shibata TF, Nishiyama T, Shigenobu S, Suzuki Y, et al. Methylome diversification through changes in DNA methyltransferase sequence specificity. *PLoS Genet* 2014;10(4):e1004272.
  24. Stoddard CI, Feng S, Campbell MG, Liu W, Wang H, Zhong X, et al. A nucleosome bridging mechanism for activation of a maintenance DNA methyltransferase. *Mol Cell* 2019;73(1):73-83.
  25. Bardina C, Colom J, Spricigo DA, Otero J, Ssnchez-Osuna M, Cortes P, et al. Genomics of three new bacteriophages useful in the biocontrol of Salmonella. *Front Microbiol* 2016;7:545.
  26. Stamereilers C, Fajardo CP, Walker JK, Mendez KN, Castro-Nallar E, Grose JH, et al. Genomic analysis of 48 Paenibacillus larvae bacteriophages. *Viruses* 2018;10(7):377.
  27. Jacobs-Sera D, Abad LA, Alvey RM, Anders KR, Aull HG, Bhalla SS, et al. Genomic diversity of bacteriophages infecting Microbacterium spp. *PLoS One* 2020;15(6):e0234636.
  28. Pawlowski A, Moilanen AM, Rissanen IA, Määttä JA, Hytonen VP, Ihalainen JA, et al. The minor capsid protein VP11 of thermophilic bacteriophage P23-77 facilitates virus assembly by using lipid-protein interactions. *J Virol* 2015;89(15):7593-603.
  29. Li X, Koc C, Kuhner P, Stierhof YD, Krismer B, Enright MC, et al. An essential role for the baseplate protein Gp45 in phage adsorption to Staphylococcus aureus. *Sci Rep* 2016;6:26455.
  30. McNulty R, Lokareddy RK, Roy A, Yang Y, Lander GC, Heck AJ, et al. Architecture of the complex formed by large and small terminase subunits from bacteriophage P22. *J Mol Biol* 2015;427(20):3285-99.
  31. Roy A, Bhardwaj A, Cingolani G. Crystallization of the nonameric small terminase subunit of bacteriophage P22. *Acta Crystallogr Sect F Struct Biol Cryst Commun* 2011;67(Pt 1):104-10.
  32. Petersen A, Josephsen J, Johnsen MG. TPW22, a lactococcal temperate phage with a site-specific integrase closely related to Streptococcus thermophilus phage integrases. *J Bacteriol* 1999;181(22):7034-42.
  33. Stoll SM, Ginsburg DS, Calos MP. Phage TP901-1 site-specific integrase functions in human cells. *J Bacteriol* 2002;184(13):3657-63.
  34. Bierbaum G, Sahl HG. Autolytic system of Staphylococcus simulans 22: influence of cationic peptides on activity of N-acetylmuramoyl-L-alanine amidase. *J Bacteriol* 1987;169(12):5452-8.
  35. LeBlanc L, Nezami S, Yost D, Tsourkas P, Amy PS. Isolation and characterization of a novel phage lysin active against Paenibacillus larvae, a honeybee pathogen. *Bacteriophage* 2015;5(4):e1080787.
  36. Malecki PH, Mitkowski P, Jagielska E, Trochimiak K, Mesnage S, Sabala I. Structural characterization of EnpA D, L-endopeptidase from Enterococcus faecalis prophage provides insights into substrate specificity of M23 peptidases. *Int J Mol Sci* 2021;22(13):7136.
  37. Ju CX, Gu HW, Lu CP. Characterization and functional analysis of atl, a novel gene encoding autolysin in Streptococcus suis. *J Bacteriol* 2012;194(6):1464-73.
  38. Kohler TP, Gisch N, Binsker U, Schlag M, Darm K, Volker U, et al. Repeating structures of the major staphylococcal autolysin are essential for the interaction with human thrombospondin 1 and vitronectin. *J Biol Chem* 2014;289(7):4070-82.
  39. Zhang L, Xu D, Huang Y, Zhu X, Rui M, Wan T, et al. Structural and functional characterization of deep-sea thermophilic bacteriophage GVE2 HNH endonuclease. *Sci Rep* 2017;7:42542.

40. Irwin DM. Evolution of the vertebrate goose-type lysozyme gene family. *BMC Evol Biol* 2014;14:188.
41. Tang F, Bossers A, Harders F, Lu C, Smith H. Complete genome sequence of the *Streptococcus suis* temperate bacteriophage  $\phi$ NJ2. *Genome Announc* 2013;1(1):e00008-12.
42. Tan Y, Zhang K, Rao X, Jin X, Huang J, Zhu J et al. Whole genome sequencing of a novel temperate bacteriophage of *P. aeruginosa*: evidence of tRNA gene mediating integration of the phage genome into the host bacterial chromosome. *Cell Microbiol* 2007;9(2):479-91.
43. O'Flaherty S, Coffey A, Edwards R, Meaney W, Fitzgerald GF, Ross RP. Genome of staphylococcal phage K: a new lineage of Myoviridae infecting gram-positive bacteria with a low G+C content. *J Bacteriol* 2004;186(9):2862-71.