Title

From phenotype to receptor: validating physiological clustering of *Escherichia coli* phages through comprehensive receptor analysis

Running title

Coliphage traits and receptor-driven classification

Keywords

Bacteriophage, *Escherichia coli*, Target receptor, Lipopolysaccharide, Phage classification

Author's names

Tomoyoshi Kaneko[a,b], Toshifumi Osaka[c], Minoru Inagaki[d], Kento Habe[a], Takuma Okabe[a,e], Satoshi Tsuneda[a,b,#]

Author's affiliations

[a] Department of Life Science and Medical Bioscience, Waseda University, Tokyo, Japan

[b] Phage Therapy Institute, Comprehensive Research Organization, Waseda University, Tokyo, Japan

[c] Department of Microbiology and Immunology, Tokyo Women's Medical University, Tokyo, Japan

[d] Department of Life Science, Faculty of Bioresources, Mie University, Mie, Japan

[e] Molecular Biosystems Research Institute, National Institute of Advanced

Industrial Science and Technology (AIST), Ibaraki, Japan

#Address correspondence corresponding author to Satoshi Tsuneda,

TWIns, 2-2 Wakamatsu-cho, Shinjuku-ku, Tokyo, Japan, 162-8480

Tel./fax: +81-3-5369-7325

E-mail: stsuneda@waseda.jp

ORCiD ID

Tomoyoshi Kaneko:

https://orcid.org/0000-0003-4857-4700

Toshifumi Osaka:

https://orcid.org/0000-0003-3628-698X

Satoshi Tsuneda:

https://orcid.org/0000-0002-4624-5359

Authorship Attribution

Tomoyoshi Kaneko: Conceptualization, Data curation, Formal analysis, Writing - original draft, Investigation

Osaka Toshifumi: Conceptualization, Supervision, Writing - review & editing

Minoru Inagaki: Conceptualization, Supervision, Writing - review & editing

Kento Habe: Data curation, Formal analysis

Takuma Okabe: Data curation, Formal analysis

Satoshi Tsuneda: Conceptualization, Supervision, Writing - review & editing

Disclosure

The authors declare no conflicts of interest associated with any financial or commercial arrangements of this manuscript.

Abstract

Understanding the relationship between bacteriophage (phage) classification and target receptors is crucial for phage ecology and applied research. In this study, we compared 13 previously isolated *Escherichia coli* phages based on physiological characteristics, whole-genome sequences, and tail fiber protein phylogenetics. We improved our previously proposed physiological clustering method by optimizing the bacterial panel for host range assessment, implementing appropriate distance metrics for mixed data types, and applying silhouette coefficient analysis for objective determination of optimal cluster numbers. We combined genomic analysis and lipopolysaccharide (LPS) structural analysis of phage-resistant *E. coli* strains to identify target receptors of the phages. Complementation experiments further confirmed the direct involvement of identified genes in phage reception. The results revealed that phylogenetically distinct *E. coli* phages target different sites in the LPS R-core region (modified by WaaV, WaaW, WaaT, and WaaY), membrane proteins (NfrB, TolA, YhaH), or flagella. Our analysis revealed that, subtle chemical modifications of LPS (such as Heptose phosphorylation) were shown to be important for *E. coli* phage recognition. Furthermore, physiological characteristics, tail fiber phylogenetics, and whole genome analysis independently classified the phages with high correlation to target receptor specificity. The addition of three phages with known receptors further validated our approach. Our results suggest that phage phenotypes and genotypes are closely related to target receptor specificity. Here, we elucidated the diversity and specificity of *E. coli* phage target receptors, providing new insights into the classification of *E. coli* phages and phage-host interactions.

Importance

Phage therapy is gaining attention as an alternative treatment for antibiotic-resistant bacteria. Developing effective phage cocktails requires combining phages with different target receptors, but traditional methods for identifying target receptors are labor-intensive. This study demonstrates that *E. coli* phages with different target receptors can be grouped based on their physiological characteristics, tail fiber sequences, or whole genomes. This insight can be used to more efficiently create diverse phage cocktails. Additionally, we identified phages targeting diverse sites, including different regions of LPS, membrane proteins, and flagella. These findings deepen our understanding of phage host recognition mechanisms, enabling the rapid preparation of effective phage cocktails and contributing to new advancements in bacterial infection treatment.

Introduction

With the emergence of antibiotic-resistant bacteria, bacteriophage (phage) therapy - utilizing viruses that specifically infect bacteria - is gaining attention as an alternative treatment. Phages exhibit extremely high host specificity, which is primarily determined by their target receptors. For instance, phages that infect *Escherichia coli* are known to target membrane proteins, such as OmpC and LamB, or the lipopolysaccharide (LPS) (1). While bacteria can also become resistant to phages, using multiple phages that target different host receptors in a therapeutic cocktail has been shown to be effective at delaying and/or suppressing the emergence of phage-resistant bacteria (2)(3)(4)(5).

However, in the clinical setting, identifying the target receptors of each phage before creating a cocktail is impractical due to time constraints. For example, in a case of multidrug-resistant *Acinetobacter baumannii* infection reported by Schooley et al., the

99 initial phage screening and selection was completed within 18 hours to address the
100 patient's critical condition, though additional phages were subsequently identified and
101 administered to the patient. In such urgent cases, conventional methods based on receptor
102 identification is impractical, and more rapid treatment decision-making is required (6).
103 Traditionally, phage target receptor identification has involved a labor-intensive process
104 of creating phage-resistant bacteria, conducting whole-genome comparative analysis, and
105 identifying mutation sites (7)(8). While this methodology is reliable, it is unsuitable for
106 clinical application due to its time and cost requirements. As such, in personalized phage
107 therapy, phages are often combined in therapeutic cocktails without the knowledge of
108 each phage's receptor. If multiple phages target the same receptor, a single receptor
109 mutation could render the cocktail ineffective.
110 In addition, while genomic approaches to phage characterization have become
111 increasingly accessible for well-studied organisms like *E. coli*, receptor identification
112 based on genomic data requires that the host bacterial genome is well-characterized, with
113 comprehensive understanding of its surface structures and their genetic determinants. For
114 emerging multidrug-resistant pathogens with less characterized genomes, such as
115 *Stenotrophomonas maltophilia* or various non-model clinical isolates, the interpretation
116 of genomic data for receptor identification remains challenging (9).
117 In our previous study, we proposed a clustering method based on physiological
118 characteristics such as growth dynamics and host range. This method makes it possible
119 to create effective phage cocktails without identifying the target receptors (10). However,
120 at the time, we did not evaluate the validity of the proposed clustering methodology.
121 First, the bacterial panel used for host range evaluation was limited and lacked systematic
122 selection criteria. Second, the distance calculation methods used for clustering did not

optimally handle the mixed nature of qualitative infection data and quantitative physiological measurements. Third, the criteria for determining the optimal number of clusters were not clearly defined. Fourth, the relationship between the clustered phages and their target receptors remained unverified. Finally, the correlation between physiological clustering and genomic characteristics was not evaluated.

The present study aimed to validate and improve the proposed clustering method through comprehensive analysis of each phage's target receptor. We enhanced our methodology by implementing more robust approaches for host range assessment, distance calculation, and cluster determination. Furthermore, we compared this improved physiological characteristic based classification method to other classification methods, including whole-genome sequencing and tail fiber protein phylogenetic analysis. Through these analyses, we aimed to evaluate how each classification method correlates with target receptors, with a particular focus on validating our refined physiological clustering approach. These findings are expected to contribute to rapid treatment decision-making in clinical settings and the standardization of phage therapy.

Results

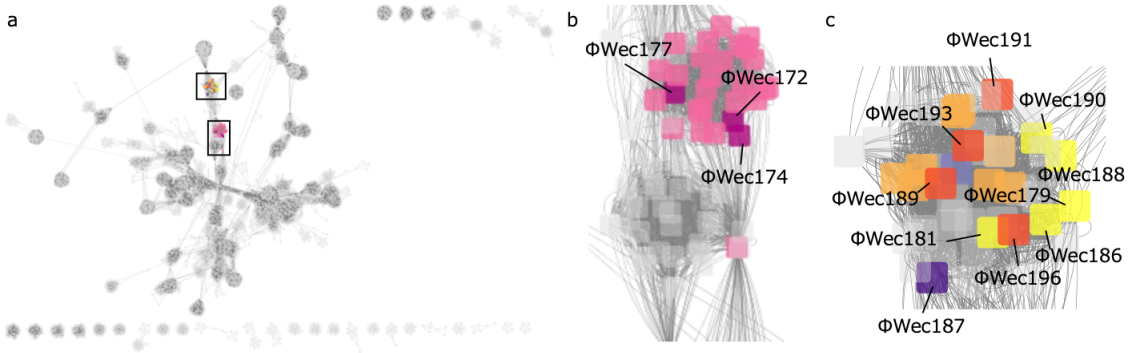
Genomic characterization of phages

Whole genome sequencing and ANI analyses indicated that the phages represent three distinct taxonomic genera (Fig. S1, Table S1, S2, S3): ΦWec172, ΦWec174, and ΦWec177 belong to *Felixounavirus* (*Ounaviridae*); ΦWec189, ΦWec191, ΦWec193, and ΦWec196 are members of *Vequintaviridae* (*Vequintavirus*); and ΦWec179, ΦWec181, ΦWec186, ΦWec187, ΦWec188, and ΦWec190 are members of *Stephanstirmvirinae*. Among the *Stephanstirmvirinae* phages, ΦWec187 showed characteristics related to

147 *Justusliebigvirus*, while our previous analysis indicated that ΦWec179, ΦWec181,

148 ΦWec186, ΦWec188, and ΦWec190 represent a proposed novel genus "*Wecvirus*" (11).

149



150

151 **Fig S1. Visualized physiological relationship.** The genomes of the 13 phages used in

152 this study and the genomes of 3503 reference phages (Prokaryotic Viral RefSeq201) from

153 vConTACT2 were clustered based on shared proteins and analyzed as a network. The

154 distance between phages was measured using vConTACT2, and the network was

155 visualized using Cytoscape. (a) Overview of the entire network showing viral clusters.

156 (b) Enlarged view of the cluster containing ΦWec172, 174, and 177. All phages in this

157 viral cluster belong to various genera within *Ounavirinae*. (c) Enlarged view showing

158 ΦWec179, 181, 186, 187, 188, and 190, which belong to or are closely related to viral

159 clusters containing phages exclusively from various genera within *Stephanstirmvirinae*

160 (11). Additionally, ΦWec189, 191, 193, and 196 belong to a viral cluster containing only

161 *Vequintavirus* phages from *Vequintavirinae*.

162

163

Table S1 ANI of phages in this study

	ΦWec187	ΦWec189	ΦWec191	ΦWec193	ΦWec196	ΦWec177	ΦWec174	ΦWec172	ΦWec179	ΦWec181	ΦWec186	ΦWec188	ΦWec190
ΦWec190									79.98	79.68	79.68	98.52	100.00
ΦWec188									79.64	79.61	79.61	100.00	99.87
ΦWec186									97.80	100.00	100.00	75.83	75.26
ΦWec181									98.41	100.00	99.34	77.09	76.50
ΦWec179									100.00	98.76	98.93	76.82	77.30
ΦWec172						99.8	99.97	100.0					
ΦWec174						99.75	100.0	99.95					
ΦWec177						100.0	99.61	99.6					
ΦWec196		87.33	84.76	99.90	100.00								
ΦWec193		85.83	84.33	100.00	99.90								
ΦWec191		87.14	100.00	84.86	85.37								
ΦWec189		100.00	89.66	87.28	87.88								
ΦWec187	100.00												

Table S2 ANI between ΦWec172, 174, 177 and registered *Ounavirinae* phages

	phiWec172	phiWec174	phiWec177	Enterobacteria_phage_Felix_01	Enterobacteria_phage_WV8	Enterobacteriaphage_UAB_Phi87	Escherichia_coli_O157_1_typing_phage_1	Escherichia_phage_EC6	Escherichia_phage_HY02	Escherichia_phage_JH2	Escherichia_phage_vB_EcoM_AIF5	Escherichia_phage_vB_EcoM_AYO145A	Escherichia_phage_vB_EcoM-Vpa81	Salmonella_phage_BPS1SQ2	Salmonella_phage_BPS17L1	Salmonella_phage_BPS17W1	Salmonella_phage_Mushroom	Salmonella_phage_Si3	Salmonella_phage_vB_SPuM_SP116	Citrobacter_phage_Michonne	Citrobacter_phage_Moogle	Citrobacter_phage_Mordin	Shigella_phage_SFI3	Shigella_phage_SFI4	Shigella_phage_SFI7	Escherichia_phage_SUSP1	Escherichia_phage_SUSP2	Staphylococcus_phage_SA
phiWec172	1.0000	0.9997	0.9881	0.7719	0.7692	0.8004	0.7751	0.7924	0.7616	0.7836	0.7698	0.7787	0.7897	0.7827	0.7909	0.8034	0.7939	0.8007	0.7551	0.5416	0.5610	0.5350	0.5491	0.5246	0.5470	0.6033	0.5985	0.7821
phiWec174	0.9995	1.0000	0.9876	0.7631	0.7597	0.8000	0.7655	0.7916	0.7520	0.7834	0.7694	0.7785	0.7893	0.7826	0.7906	0.8031	0.7939	0.8004	0.7547	0.5414	0.5606	0.5348	0.5489	0.5244	0.5468	0.6031	0.5982	0.7818
phiWec177	0.9881	0.9878	1.0000	0.7507	0.7676	0.7628	0.7687	0.7758	0.7526	0.7746	0.7365	0.7745	0.7758	0.7457	0.7799	0.7680	0.7822	0.7689	0.7833	0.5387	0.5480	0.5321	0.5277	0.5227	0.5200	0.5671	0.5659	0.7541
Enterobacteria_phage_Felix_01	0.7756	0.7733	0.7756	1.0000	0.7684	0.8177	0.7840	0.7681	0.7747	0.7865	0.7921	0.7910	0.8018	0.8074	0.7328	0.7611	0.8426	0.8297	0.7746	0.5260	0.5481	0.5379	0.5469	0.5570	0.5453	0.5803	0.5896	0.7999
Enterobacteria_phage_WV8	0.7481	0.7400	0.7569	0.7714	1.0000	0.7974	0.9496	0.7931	0.8383	0.7707	0.7945	0.8372	0.8315	0.8253	0.7502	0.7335	0.7855	0.7152	0.7414	0.5296	0.5599	0.5236	0.5656	0.5587	0.5696	0.5621	0.5825	0.8164
Enterobacteriaphage_UAB_Phi87	0.7529	0.7444	0.7529	0.8116	0.8157	1.0000	0.8043	0.8026	0.8096	0.8039	0.8142	0.8531	0.8223	0.8150	0.7884	0.7940	0.8341	0.7666	0.7887	0.5254	0.5225	0.5354	0.5306	0.5196	0.5387	0.5896	0.6253	0.8445
Escherichia_coli_O157_typing_phage_1	0.7052	0.6878	0.6969	0.7769	0.6364	0.8121	1.0000	0.7699	0.8069	0.7489	0.8180	0.8064	0.8085	0.7909	0.7279	0.6997	0.7411	0.6978	0.7417	0.4955	0.5129	0.5011	0.5276	0.5139	0.5376	0.5609	0.5935	0.8012
Escherichia_phage_EC6	0.7801	0.7800	0.7805	0.7794	0.7973	0.8239	0.7940	1.0000	0.8205	0.7879	0.7630	0.8178	0.8313	0.7876	0.7289	0.7618	0.7719	0.7441	0.7803	0.5379	0.5392	0.5373	0.5707	0.5678	0.5610	0.5729	0.5631	0.7995
Escherichia_phage_HY02	0.7534	0.7439	0.7512	0.7606	0.8095	0.7676	0.8049	0.7936	1.0000	0.7493	0.7632	0.8107	0.8033	0.7926	0.7218	0.7321	0.7863	0.6994	0.7372	0.4926	0.5287	0.4904	0.5354	0.5426	0.5275	0.5599	0.5601	0.7740
Escherichia_phage_JH2	0.7631	0.7630	0.7631	0.7507	0.7641	0.8134	0.7829	0.7537	0.7434	1.0000	0.7522	0.7724	0.7745	0.8188	0.7596	0.8153	0.7998	0.7448	0.7666	0.5587	0.5691	0.5610	0.5712	0.5572	0.5814	0.6214	0.6094	0.8117
Escherichia_phage_vB_EcoM_AIF5	0.7509	0.7427	0.7509	0.7912	0.8258	0.8359	0.8281	0.7735	0.7908	0.7803	1.0000	0.8230	0.8141	0.7745	0.7759	0.7524	0.8174	0.7590	0.7790	0.5375	0.5772	0.5464	0.5579	0.5548	0.5729	0.5882	0.6158	0.7926
Escherichia_phage_vB_EcoM_AYO145A	0.7522	0.7458	0.7574	0.7797	0.8165	0.8462	0.8063	0.7965	0.8132	0.7851	0.8155	1.0000	0.8085	0.8326	0.7285	0.7351	0.8083	0.7332	0.7851	0.5340	0.5427	0.5272	0.5270	0.5306	0.5207	0.6036	0.6165	0.8231
Escherichia_phage_vB_EcoM-Vpa81	0.7597	0.7508	0.7595	0.7885	0.8253	0.8003	0.8237	0.7929	0.8144	0.7874	0.8077	0.8123	1.0000	0.8053	0.7449	0.7170	0.8023	0.7546	0.7399	0.5025	0.5494	0.5246	0.5236	0.5397	0.5357	0.5946	0.5891	0.7790
Salmonella_phage_BPS1SQ2	0.7283	0.7283	0.7203	0.7836	0.8162	0.7900	0.7965	0.7550	0.7812	0.7803	0.7460	0.8256	0.8207	1.0000	0.7452	0.7659	0.8038	0.7360	0.7343	0.5134	0.5511	0.5058	0.5211	0.5163	0.5320	0.5536	0.5482	0.8320
Salmonella_phage_BPS17L1	0.8008	0.7921	0.7923	0.7724	0.7959	0.8153	0.7805	0.7787	0.7728	0.8492	0.8068	0.8049	0.7908	0.8253	1.0000	0.8220	0.8187	0.8192	0.8114	0.5470	0.5347	0.5430	0.5647	0.5392	0.5439	0.5872	0.6033	0.7905
Salmonella_phage_BPS17W1	0.7833	0.7833	0.7837	0.7422	0.7470	0.7988	0.7520	0.7552	0.7474	0.8325	0.7550	0.7492	0.7705	0.8196	0.8003	1.0000	0.8175	0.7621	0.7663	0.5128	0.5286	0.5152	0.5449	0.5221	0.5270	0.5646	0.5659	0.8002
Salmonella_phage_Mushroom	0.7707	0.7707	0.7767	0.8196	0.7840	0.8160	0.7654	0.7637	0.7959	0.8106	0.8013	0.8117	0.8153	0.8361	0.7764	0.8047	1.0000	0.8399	0.7664	0.5370	0.5516	0.5320	0.5343	0.5221	0.5386	0.5996	0.5820	0.8067
Salmonella_phage_Si3	0.7922	0.7856	0.7835	0.8392	0.7533	0.8125	0.7628	0.7431	0.7230	0.8136	0.7949	0.7645	0.7902	0.8137	0.8227	0.7639	0.8882	1.0000	0.7509	0.5289	0.5494	0.5220	0.5381	0.5311	0.5478	0.5903	0.5792	0.7775
Salmonella_phage_vB_SPuM_SP116	0.7440	0.7440	0.7359	0.7489	0.7631	0.8235	0.7669	0.7681	0.7597	0.7896	0.7624	0.8025	0.7262	0.7820	0.7549	0.7746	0.7678	0.7132	1.0000	0.5194	0.5010	0.5193	0.5202	0.5100	0.5177	0.5546	0.5869	0.8049
Citrobacter_phage_Michonne	0.5235	0.5159	0.5240	0.5180	0.5042	0.5465	0.5260	0.5262	0.5013	0.5276	0.4976	0.5275	0.5124	0.5082	0.5137	0.5173	0.5159	0.4891	0.5148	1.0000	0.7677	0.9459	0.7374	0.7611	0.7792	0.6875	0.5987	0.5484
Citrobacter_phage_Moogle	0.5393	0.5393	0.5393	0.5428	0.5698	0.5469	0.5748	0.5392	0.5529	0.5393	0.5468	0.5555	0.5701	0.5617	0.4909	0.5226	0.5577	0.5297	0.5043	0.7920	1.0000	0.7859	0.7500	0.7685	0.7663	0.7314	0.6550	0.5525
Citrobacter_phage_Mordin	0.5141	0.5082	0.5141	0.5408	0.5411	0.5529	0.5568	0.5254	0.5234	0.5524	0.5381	0.5487	0.5438	0.5502	0.5335	0.5260	0.5412	0.5146	0.5124	0.9258	0.7873	1.0000	0.7567	0.7900	0.8085	0.6875	0.6141	0.5593
Shigella_phage_SFI3	0.5523	0.5445	0.5523	0.5393	0.5721	0.5624	0.5650	0.5432	0.5597	0.5532	0.5610	0.5717	0.5478	0.5332	0.5658	0.5360	0.5429	0.5313	0.5466	0.7590	0.7259	0.7589	1.0000	0.8446	0.9031	0.6890	0.6352	0.5512
Shigella_phage_SFI4	0.4909	0.4854	0.4909	0.5082	0.5359	0.5357	0.5233	0.5477	0.5258	0.5135	0.5066	0.5427	0.5194	0.5144	0.4908	0.5009	0.5062	0.4935	0.4975	0.7911	0.7685	0.7969	0.8719	1.0000	0.8703	0.6673	0.6216	0.5210
Shigella_phage_SFI7	0.4945	0.4945	0.4945	0.5409	0.5694	0.5552	0.5796	0.5352	0.5348	0.5462	0.5406	0.5745	0.5416	0.5719	0.4977	0.5313	0.5657	0.5173	0.5220	0.8028	0.7673	0.8082	0.8753	0.8651	1.0000	0.6897	0.6515	0.5726
Escherichia_phage_SUSP1	0.5369	0.5314	0.5369	0.5407	0.5396	0.5334	0.5221	0.5237	0.5245	0.5478	0.5405	0.5427	0.5403	0.5513	0.5090	0.5228	0.5591	0.5258	0.4950	0.6589	0.6617	0.6641	0.6520	0.6320	0.6685	1.0000	0.8047	0.5926
Escherichia_phage_SUSP2	0.5655	0.5654	0.5655	0.5575	0.5681	0.5617	0.5549	0.5625	0.5741	0.5539	0.5687	0.6031	0.5603	0.5541	0.5508	0.5613	0.5622	0.5331	0.5671	0.6000	0.6153	0.6071	0.6280	0.6136	0.6296	0.8243	1.0000	0.5622
Staphylococcus_phage_SA	0.4417	0.4416	0.4417	0.4804	0.4824	0.4945	0.4834	0.4496	0.4737	0.4791	0.4537	0.4835	0.4718	0.5019	0.4409	0.4640	0.4764	0.4348	0.4614	0.3162	0.3075	0.3244	0.3284	0.3183	0.3229	0.3493	0.3568	1.0000

The table shows ΦWec172, 174, and 177 in the upper left. Maximum ANI values with registered phages range between 70% and 80%, suggesting these phages represent new species within the *Felixounavirus* genus of *Ounavirinae*, considering the established thresholds of 70% for genus demarcation and 95% for species demarcation. For detailed data, refer to the supplementary information Excel file.

Table S3 ANI between ΦWec189, 191, 193, 196 and registered *Vequintavirus* phages

	phiWec189	phiWec191	phiWec193	phiWec196	Escherichia~phage~APECc02	Escherichia~phage~FFH2	Escherichia~phage~FV3	Escherichia~phage~JES2013	Escherichia~phage~Murica	Escherichia~phage~slur16	Escherichia~phage~V18	Escherichia~phage~V5
phiWec189	**1.0000**	**0.8713**	**0.8582**	**0.8731**	**0.8559**	**0.8715**	**0.8257**	**0.8797**	**0.8950**	**0.8814**	**0.8452**	**0.8684**
phiWec191	**0.8956**	**1.0000**	**0.8431**	**0.8475**	**0.8882**	**0.8781**	**0.8308**	**0.8719**	**0.9152**	**0.8930**	**0.8312**	**0.8702**
phiWec193	**0.8725**	**0.8485**	**1.0000**	**0.9980**	**0.8711**	**0.8845**	**0.7927**	**0.8688**	**0.8983**	**0.8490**	**0.8152**	**0.8684**
phiWec196	**0.8785**	**0.8536**	**0.9980**	**1.0000**	**0.8641**	**0.8839**	**0.7986**	**0.8628**	**0.8979**	**0.8488**	**0.8095**	**0.8740**
Escherichia~phage~APECc02	**0.8634**	**0.8558**	**0.8552**	**0.8627**	1.0000	0.8921	0.8060	0.8489	0.8772	0.8503	0.8185	0.8450
Escherichia~phage~FFH2	**0.8891**	**0.8724**	**0.8603**	**0.8827**	0.9084	1.0000	0.8491	0.9051	0.9085	0.8801	0.8529	0.8636
Escherichia~phage~FV3	**0.8359**	**0.8292**	**0.8137**	**0.8297**	0.8019	0.8451	1.0000	0.8360	0.8477	0.8629	0.8243	0.8204
Escherichia~phage~JES2013	**0.8640**	**0.8540**	**0.8330**	**0.8633**	0.8515	0.8906	0.8336	1.0000	0.9124	0.8766	0.8792	0.8931
Escherichia~phage~Murica	**0.8884**	**0.8994**	**0.8710**	**0.8899**	0.8722	0.8991	0.8186	0.8991	1.0000	0.8898	0.8551	0.8600
Escherichia~phage~slur16	**0.8789**	**0.8951**	**0.8528**	**0.8578**	0.8610	0.8860	0.8670	0.8817	0.9090	1.0000	0.8736	0.8749
Escherichia~phage~V18	**0.8112**	**0.7778**	**0.7596**	**0.7824**	0.7854	0.8042	0.7813	0.8216	0.8428	0.8145	1.0000	0.7883
Escherichia~phage~V5	**0.8889**	**0.8699**	**0.8486**	**0.8642**	0.8537	0.8531	0.8274	0.8841	0.8787	0.8761	0.8542	1.0000

The table shows ΦWec189, 191, 193, and 196 in the upper left. Maximum ANI values with registered phages range between 70% and 90%, suggesting these phages represent new species within the *Vequintavirus* genus, considering the established thresholds of 70% for genus demarcation and 95% for species demarcation. For detailed data, refer to the supplementary information Excel file.

Whole genome analysis of TK001 wild type strain and mutation analysis of target receptor mutant strains

TK001 used in this study was isolated from the feces of a C57BL/6JJcl female mouse with dextran sodium sulfate induced colitis (MP Biomedicals, Soho, OH, USA) and was used as the host strain for phage isolation. Hybrid assembly of shotgun sequencing and long-read sequencing revealed that the complete genome of TK001 consists of 4,779,140 bp with a GC content of 50.79%, and its sequence type (ST) based on Multilocus sequence typing (MLST) analysis is 345.

Mutation analysis of 13 phage-resistant strains revealed frameshift mutations or stop codon insertions in genes encoding outer membrane/inner membrane proteins or LPS synthesis enzymes in each strain (Table 1). Among the 13 phage-resistant strains, 8 strains showed mutations in the genes of the LPS synthesis pathway, while the other 4 strains

196 had mutations in genes related to membrane proteins or flagella. Notably, no target
197 receptor mutants were obtained for ΦWec179. For ΦWec177, two resistant strains were
198 analyzed and are distinguished as R177-1 and R177-2.
199

200 **Table 1 Genes with frameshift or stop codon mutations in target receptor mutant**
201 ***E. coli* strains**

phage used to make strains resistant	strain	mutated gene(s)	description (role of genes)	mutation site in WT	mutation type	mutation site in WT amino acids (native length [AAs])	the size of the expected gene product (WT size) [Da]
ΦWec172	R172	waaV	beta1,3-glucosyltransferase	3934566	Frame shift (A→.)	p.Y151Ifs*158 (327)	17856.774 (38773.742)
ΦWec174	R174	waaW	UDP-galactose--(galactosyl) LPS alpha1,2-galactosyltransferase	3935258	Frame shift (.→T)	p.A290Dfs*295 (341)	33723.99 (39312.395)
ΦWec177	R177-1	waaV	beta1,3-glucosyltransferase	3934492	Frame shift (A→.)	p.F175Lfs*181 (327)	20786.28 (38773.742)
ΦWec177	R177-2	waaV	beta1,3-glucosyltransferase	3934492	Frame shift (A→.)	p.F175Lfs*181 (327)	20786.28 (38773.742)
		wecA	Undecaprenyl-phosphate alpha-*N*-acetylglucosaminyl 1-phosphate transferase	4108976	Frame shift (A→.)	p.A28Afs*28 (367)	3647.418 (40938.761)
ΦWec181	R181	manB-1	Mannose-1-phosphate guanylyltransferase	2178550-2178553	Frame shift (CCTC→.)	p.A390_E391fs*432 (458)	47602.579 (50839.022)
		waaT	LPS 1,2-glucosyltransferase	3937043	Stop codon (C→T)	p.W269* (331)	30982.807 (38497.428)
ΦWec186	R186	fihD	Flagellar transcriptional regulator FlhD	2021475	Ser→Pro (A→G)	p.S16P (156)	13308.283 (13298.245)
		bcsG	Cellulose biosynthesis protein BcsG	3831107	Stop codon (C→T)	p.Q542* (559)	59871.85 (61986.166)
ΦWec187	R187	bcsG	Cellulose biosynthesis protein BcsG	3831107	Stop codon (C→T)	p.Q542* (559)	59871.85 (61986.166)
ΦWec188	R188	yhaH	Inner membrane protein YhaH	3405873-3405920	Frame shift*	p.Pro66_Ala82del (121)	12236.536 (14262.756)
ΦWec189	R189	nfrB	bacteriophage N4 adsorption protein B	576511	Stop codon (C→T)	p.W391* (745)	45051.426 (85329.404)
		ompA	Outer membrane protein A	1046632-1046640	Frame shift (CACCAAAAG→.)	p.A60_G63del (346)	36906.975 (37182.278)
ΦWec190	R190	waaG	UDP-glucose:(heptosyl)LPS alpha-1,3-glucosyltransferase	3940586	Stop codon (A→T)	p.L76*(374)	8796.029 (42509.807)
ΦWec191	R191	nfrB	bacteriophage N4 adsorption protein B	576613	Asp→Ala (T→G)	p.N357A (745)	85285.395 (85329.404)
		tolA	TolA	767756	Stop codon (C→T)	p.Q269* (422)	7295.244 (43138.074)
ΦWec193	R193	waaY	LPS core heptose(II) kinase RfaY	3936259	Stop codon (C→A)	p.E196* (230)	22793.077 (26975.989)
ΦWec196	R196	waaY	LPS core heptose(II) kinase RfaY	3936259	Stop codon (C→A)	p.E196* (230)	22793.077 (26975.989)

202 In the "Mutation type" column, an asterisk (*) indicates that the sequence "CGTGGTGGGCAGTTCAGTTCCGCCGCCTGCACGACACCGACCGCTCGG" is missing (i.e., replaced with ".").
The "Mutation site in WT amino acids (native length [AAs])" column follows HGVS nomenclature.

203

204 **Identification and validation of phage receptor genes through efficiency of plating**
205 **(EoP) of phages against target receptor mutant strains and Keio collection,**
206 **complementation, and structural prediction**

207 The EoP of each phage against TK001 and target receptor mutant strains was evaluated
208 (Table 2). A decrease in EoP to less than 10-fold ($\log(\mathrm{EoP}) < -1$) was defined as a
209 reduction in infection efficiency. All phages could infect *bcsG* mutant (R187), a mutation
210 common to R186 (*fihD* and *bcsG* mutant); therefore, *bcsG* was not considered to be
211 involved in target recognition.
212 Regarding LPS synthesis genes, none of the phages could infect the *waaG* mutant (R190).
213 ΦWec172, 174, and 177 lost their ability to infect *waaW* mutant (R174) and *waaV*
214 mutants (R172, R177-1, and R177-2). ΦWec174, 177, 187, 189, 191, 193, and 196 could

215 not infect *waaY* mutants (R193 and R196). ΦWec179, 181, 186, 188, and 190 lost
216 infectivity against *waaT* and *manB-1* double mutant (R181).
217 For membrane protein and flagella genes, ΦWec179, 181, and 186 could not infect *fihD*
218 mutant (R186). ΦWec188 and 190 lost infectivity against *yhaH* mutant (R188). ΦWec189,
219 191, 193, and 196 could not infect *nfrB* and *ompA* double mutant (R189) and *nfrB* and
220 *tolA* double mutant (R191).
221 Using ΦWec187, 189, 191, 193, and 196, which can infect the K-12-derived BW25113
222 strain (wild-type of the Keio collection), we investigated their ability to infect Keio
223 collection strains with single gene deletions corresponding to the mutations found in
224 phage-resistant strains (Table S4). ΦWec187, 189, 191, 193, and 196 all failed to infect
225 Δ*nfrB*, Δ*waaG*, and Δ*waaF* strains. Additionally, ΦWec189, 191, 193, and 196 could not
226 infect the Δ*nfrA* strain, which is genetically redundant with *nfrB*.
227 Based on these results, the genes necessary for infection by each phage were classified
228 into six types (Table 3). Types a and b were very similar, differing only in the presence or
229 absence of a *waaY* mutation.
230 To further validate our receptor identification, we performed two additional analyses.
231 First, we tested three additional phages (ΦWec270, ΦWec272, and T5) against our panel
232 of receptor mutant strains. ΦWec270 and ΦWec272, which belong to the same genus as
233 ΦWec189, 191, 193, and 196 based on previous genomic analyses, showed identical
234 infection patterns to these phages, confirming they share the same target receptor
235 requirements including both NfrB and phosphorylated heptose in LPS (mediated by
236 WaaY). In contrast, T5, which is known to target FhuA in *E. coli* K-12 (12)(13) and is
237 taxonomically distinct from all other phages used in this study, remained infective against
238 all our receptor mutant strains, indicating it recognizes a different receptor not affected

239 by our mutations.

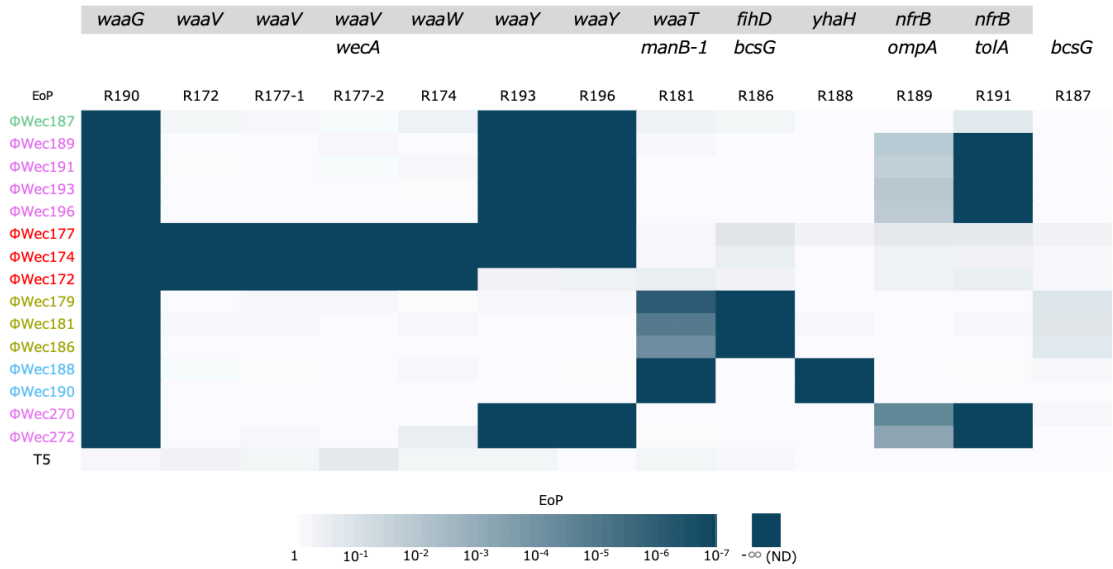
240 Second, we conducted complementation experiments by expressing wild-type genes in
241 the corresponding mutant strains using plasmid vectors. When the wild-type versions of
242 the mutated genes were reintroduced, infection sensitivity was restored in most cases,
243 confirming the direct involvement of these genes in phage reception (Table 3).
244 Specifically, complementation of *waaW*, *waaG*, *waaT*, *waaY*, *nfrB*, *flhD*, and *yhaH*
245 mutations restored sensitivity to the respective phages that lost infectivity against these
246 mutants.

247 Interestingly, complementation results for *waaV* mutations were variable. Introduction of
248 wild-type *waaV* into the R172 mutant (with deletion at position 451 bp) restored phage
249 sensitivity, while complementation failed to restore sensitivity in R177-1 and R177-2
250 mutants (both with deletion at position 525 bp). To investigate whether this was due to
251 the mutant protein interfering with wild-type function or due to polar effects on
252 downstream genes, we expressed the mutant *waaV* variants in wild-type TK001. Neither
253 variant affected EoP, suggesting that the different complementation outcomes might be
254 related to secondary effects specific to the R177-1/2 genetic background.

255 Structural predictions of LPS synthesis enzymes using AlphaFold2 and AlphaFold3
256 revealed the molecular impacts of the identified mutations on protein structure and
257 function (detailed analyses in Supplementary Results and Discussion, Fig. S3-9). The
258 analysis revealed that mutations in *waaG*, *waaV*, *wecA*, and *manB-1* resulted in extensive
259 truncations (45-92% of the protein). Mutations in *waaW*, *waaT*, and *waaY* resulted in
260 more limited truncations (14-19% of the protein), primarily affecting C-terminal regions
261 rich in basic amino acids while preserving most catalytic residues.

262

263 **Table 2 Infectivity of each phage against phage-resistant *E. coli* strains**

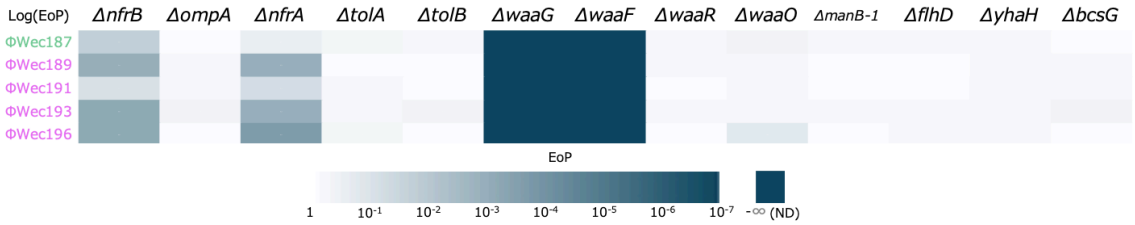


Note: Gray indicates LPS genes. For ΦWec179, 181, and 186 against R181, 4 out of 7 trials were ND(Not Detected), and the log (EoP) value in the table is the average of the remaining 3 trials. Additionally, phage names with the same color indicate groups that show similar EoP patterns against the panel of phage-resistant mutant strains, reflecting similar receptor specificity profiles.

264

265

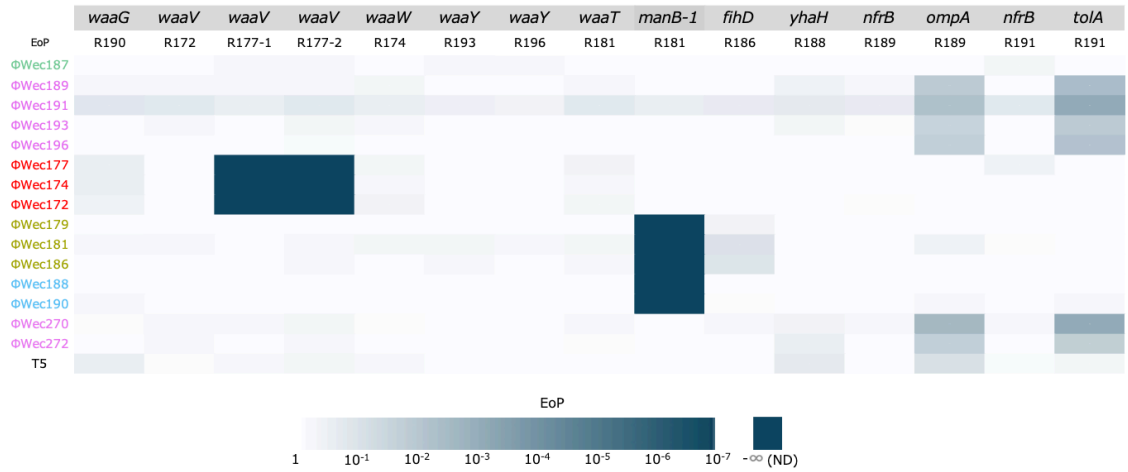
266 **Table S4 Infectivity of K-12-infecting phages against Keio collection strains**



267 Note: Phage names of the same color indicate that they belong to the same group when divided into five sub-groups in that grouping.

268

269 **Table 3 Infectivity of each phage against complemented resistant mutants**



270 Note: The order and color coding of phages is maintained consistent with Table 2 for ease of comparison. Phages of the same color show similar infectivity patterns against both the resistant mutants (Table 2) and their complemented derivatives.

271

272 **Table 4 Genes associated with molecules recognized by each phage**

	waaV	*waaW*	*waaY*	*waaT*	*fihD*	*yhaH*	*nfrB*	*tolA*	Type
ΦWec172	1	1	0	0	0	0	0	0	a
ΦWec174	1	1	1	0	0	0	0	0	b
ΦWec177	1	1	1	0	0	0	0	0	
ΦWec179	0	0	0	1	1	0	0	0	c
ΦWec181	0	0	0	1	1	0	0	0	
ΦWec186	0	0	0	1	1	0	0	0	
ΦWec187	0	0	1	0	0	0	0	0	d
ΦWec188	0	0	0	1	0	1	0	0	e
ΦWec190	0	0	0	1	0	1	0	0	
ΦWec189	0	0	1	0	0	0	1	1	f
ΦWec191	0	0	1	0	0	0	1	1	
ΦWec193	0	0	1	0	0	0	1	1	
ΦWec196	0	0	1	0	0	0	1	1	
ΦWec270	0	0	1	0	0	0	1	1	
ΦWec272	0	0	1	0	0	0	1	1	
T5	0	0	0	0	0	0	0	0	g

273 Note: 1 indicates "targeting", 0 indicates "not targeting". LPS is represented by synthesis genes.

274

275 **LPS structure of TK001 wild type and phage resistant strains**

276 Mutations in genes involved in the LPS synthesis pathway were identified in many phage-
277 resistant *E. coli* strains. Most of the LPS-related enzymes mutated in the target receptor
278 mutant strains were associated with R-core synthesis, while *wecA* and *manB-1* were genes
279 for synthesizing sugars that constitute the O-antigen. Each *E. coli* strain possesses one of

280 five R-core types, and the analysis suggested that the R-core of TK001 was the R1 type
281 (Fig. 1a, Fig. S2a) (14). Variations in the R-core structure were suggested to arise from
282 deficiencies in the R-core synthesis pathway genes (Fig. S2b). The KEGG mapper results
283 indicated the presence of both *waaR* and *waaT* genes, as these genes share the same
284 KEGG Orthology (KO) identifier in the database, though they encode functionally
285 distinct enzymes. However, examination of the actual genome annotation revealed that
286 TK001 possesses only the *waaT* gene and not *waaR*, suggesting that the cascade beyond
287 *waaR* is not active.
288 DOC-PAGE analysis of LPS structure revealed that TK001 is similar to *E. coli* O111:B4,
289 suggesting the presence of LPS with a relatively uniform, long O-antigen repeat unit in
290 addition to the R-core (Fig. 1b). Strains with mutations in genes unrelated to LPS
291 synthesis showed the same band pattern in DOC-PAGE as TK001, indicating no change
292 in LPS structure (Fig. 1c). In *waaV* mutants (R172, R177-1, and R177-2), O-antigen (the
293 upper band in DOC-PAGE) completely disappeared. The DOC-PAGE profile of the
294 *waaW* mutant (R174) showed changes in the shape of the upper band and a downward
295 shift in the R-core band position. The *waaT* mutant (R181) showed complete
296 disappearance of the upper band and a downward shift in the R-core band position. The
297 *waaG* mutant (R190) showed near disappearance of the upper band and separation of the
298 R-core band into two. No clear changes in band pattern were observed in *waaY* mutants
299 (R193 and R196).

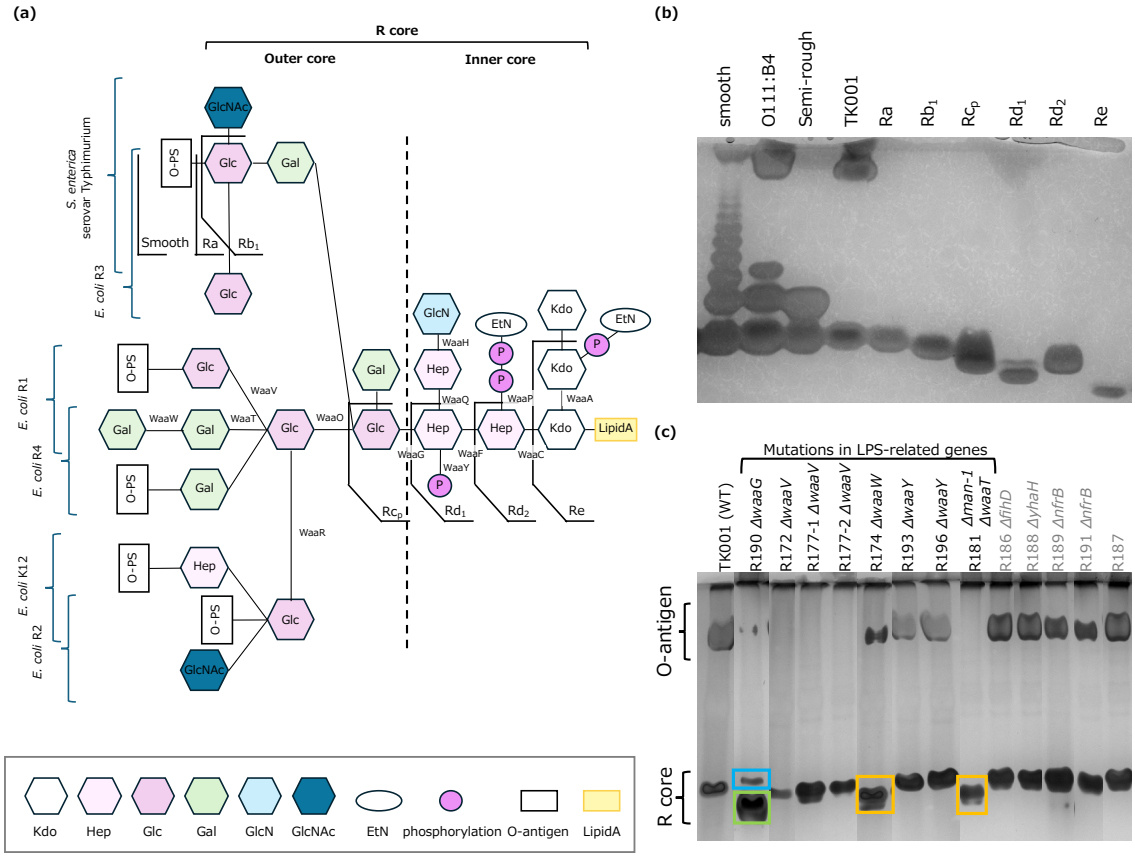
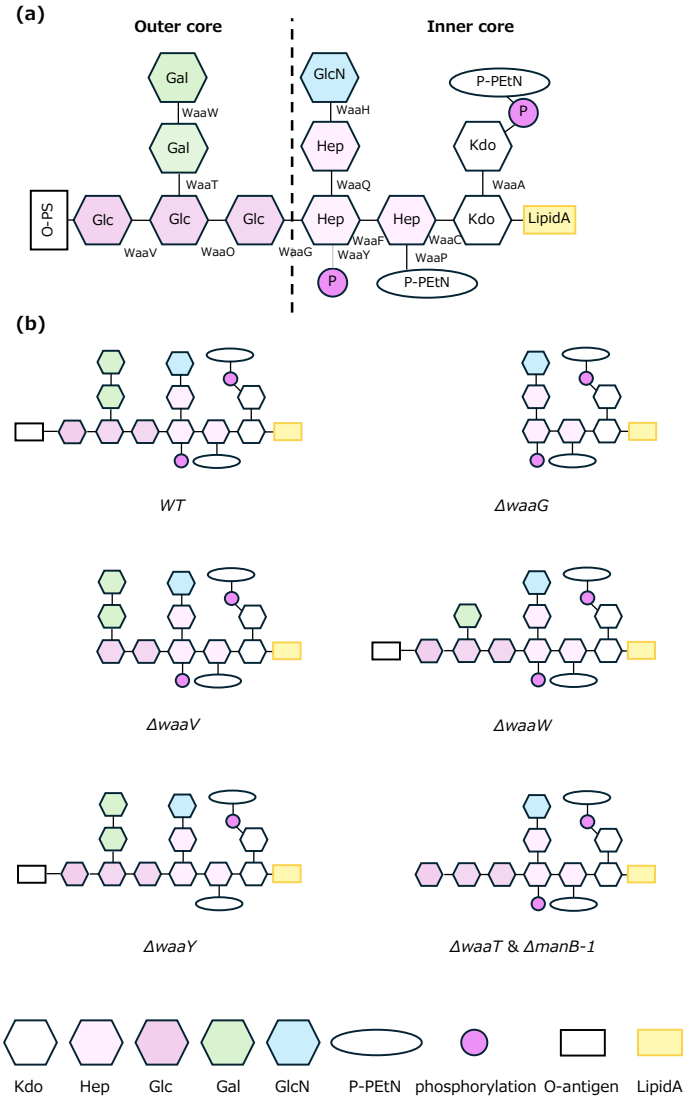


Fig. 1 LPS structural analysis and comparison across reference strains and TK001 mutants

(a) Schematic representations of LPS structures of *Salmonella enterica* serovar Typhimurium and *E. coli* R1, R2, R3, R4 and K12. The predicted LPS structure of TK001 and associated biosynthesis enzymes is also shown. (b) DOC-PAGE analysis comparing LPS from TK001 with reference strains. "Smooth" is LPS from *S. enterica* serovar Typhimurium, showing ladder-like bands characteristic of smooth strains with varying numbers of repeats of O-antigen. "O111:B4" is LPS from *Escherichia coli* O111:B4, showing a band pattern characteristic of smooth strains with relatively compact long repeat units. "Semi-rough" is LPS from *S. enterica* serovar Typhimurium SL1034 (semi-rough chemotype). “Ra” is Ra chemotype. Same as others. (c) DOC-PAGE analysis of LPS from TK001 phage-resistant mutant strains. The order has been rearranged for ease

313 of viewing. Orange and green highlights show the downward shift of the lower bands
314 compared to the wild type. Blue highlights indicate the presence of an additional band
315 distinct from the band marked by the green highlight.
316



317

318 **Fig. S2 Prediction of LPS synthesis pathway in *Escherichia coli* TK001 strain by**
319 **KEGG mapping**

320 (a) Structure of R1-type R-core and related synthesis enzymes. (b) Predicted LPS
321 structures of LPS-related gene mutants.

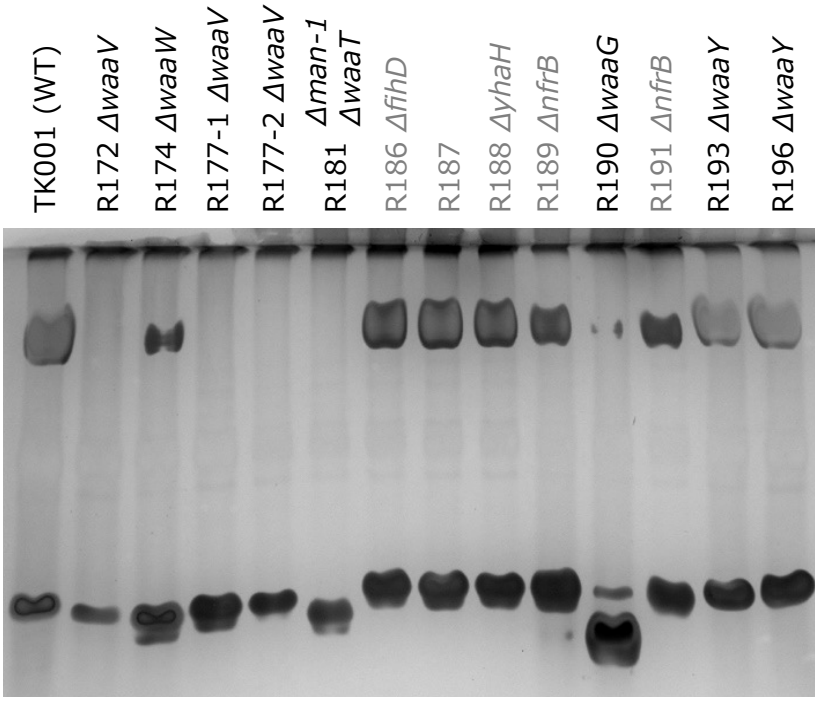


Fig. S3 Original unedited DOC-PAGE gel images of LPS from wild-type TK001 and phage-resistant mutant strains

Original, unedited DOC-PAGE gel image showing LPS from wild-type TK001 and all phage-resistant mutant strains. All samples were run simultaneously on the same gel under identical conditions, ensuring accurate size comparisons between samples. In Fig. 1(c) of the main text, lanes from this gel have been rearranged for ease of viewing and logical comparison, but no other adjustments were made to the image that would affect the relative positions or intensities of the bands.

Improved methodology for physiological clustering of phages

To enhance the rigor of our previously proposed classifiying method based on physiological characteristics, we implemented several methodological improvements.

First, we optimized the bacterial panel for host range assessment by analyzing an infection matrix of 58 phages, which were primarily those used in our previous research (15), against 41 *Enterobacteriaceae* strains (Table S5, 6). Using a greedy algorithm, we selected strains that collectively maximized the number of distinct infection patterns observed. This approach identified 16 strains (Table 5) that collectively revealed 39 distinct infection patterns, providing a more systematic basis for host range evaluation than previously used.

For clustering analyses, we implemented Gower distance instead of Euclidean distance to appropriately handle both qualitative infection data and quantitative physiological measurements. To objectively determine the optimal number of clusters, we applied silhouette coefficient analysis, which identified k=4 as the optimal number of clusters for our original set of 13 phages (Fig. 2a).

The resulting clustering (Fig. 2d) revealed four distinct phage groups: P1 (ΦWec189, 191, 193, 196), P2 (ΦWec172, 174, 177), P3 (ΦWec179, 181, 186), and P4 (ΦWec187, 188, 190). When additional phages with known receptors (ΦWec270, ΦWec272, and T5) were included, silhouette analysis identified k=5 as the optimal number, with T5 forming its own distinct group (P5), while ΦWec270 and ΦWec272 clustered with group P1 (Fig 2a, g).

355 **Table 5 Bacterial panel selected for optimized host range assessment**

Rank	*Enterobacteriaceae* strain	Additional discriminatory patterns	Cumulative discriminatory patterns
1	TK001	2	2
2	MG1655	2	4
3	BL21	3	7
4	ESBL953	5	12
5	NBRC102203	6	18
6	ESBL983	4	22
7	ATCC43888	3	25
8	TOP10F′	2	27
9	ESBL946	2	29
10	ESBL955	2	31
11	ESBL1054	2	33
12	*E. fergusonii* NBRC102419	2	35
13	SUTL-3	1	36
14	ESBL920	1	37
15	ESBL1013	1	38
16	*S*. Pullorum NBRC3163	1	39

356 Note: All of them are *E. coli* other than *E. ferugusonii* and *S*. Pullorum

357

358

359 **Table S5 Bacterial strains tested for optimization of host range assessment panel**

Bacteria	Origin/characteristics
E. coli TK001 (strain used to isolate phage)	Derived from colitis induced mouse feces in the SPF room
MG1655	Laboratory strain
JM109	
TY0807	
TOP10F'	
BW25113	
BL21	
SUTL-1	Blood culture isolate: ExPEC, ESBL-producing
SUTL-2	
SUTL-3	
SUTL-4	
SUTL-5	
ESBL920	Mid-stream urine isolate, ESBL-producing
ESBL922	Venous blood isolate, ESBL-producing
ESBL933	Venous blood isolate, ESBL-producing
ESBL946	Venous blood isolate, ESBL-producing
ESBL953	Venous blood isolate, ESBL-producing
ESBL955	Pus isolate, ESBL-producing
ESBL960	Mid-stream urine isolate, ESBL-producing
ESBL963	Pus isolate, ESBL-producing
ESBL971	Pus isolate, ESBL-producing
ESBL981	Catheter urine isolate, ESBL-producing
ESBL983	Venous blood isolate, ESBL-producing
ESBL991	Venous blood isolate, ESBL-producing
ESBL994	Venous blood isolate, ESBL-producing
ESBL1002	Venous blood isolate, ESBL-producing
ESBL1013	Catheter urine isolate, ESBL-producing
ESBL1054	Perianal ulcer isolate, ESBL-producing
ESBL1064	Mid-stream urine isolate, ESBL-producing
ESBL1080	Bladder catheter urine isolate, ESBL-producing
ESBL1090	Pus isolate, ESBL-producing
ESBL1097	Decubitus ulcer isolate, ESBL-producing"
ESBL1109	Mid-stream urine isolate, ESBL-producing
ATCC43888	Purchased from ATCC
NBRC102203	Purchased from NBRC
TWCC59747	Positive blood culture strains from patients with suspected Bacteria translocation at Tokyo Women's Medical University
E. fergusonii NBRC102419	Purchased from NBRC
S. pullorum NBRC3163	Purchased from NBRC
S. typhimurium NBRC13245	Purchased from NBRC
S. enteritidis NBRC3313	Purchased from NBRC
S. Minnesota NBRC15335	Purchased from NBRC

360

Table S6 Infection matrix showing host range patterns of phages against bacterial strains

Note: Phages with the same phylogenetic classification are indicated by the same color. A value of 1 indicates infection (confirmed by plaque formation), while 0 indicates no infection.

Comparison of target receptors with various classification methods

To evaluate how our improved physiological clustering correlates with target receptor specificity, we compared groupings based on physiological characteristics, tail fiber phylogeny, and whole genome phylogeny with grouping based on target receptors. Using silhouette coefficient analysis, we determined the optimal cluster numbers for tail fiber phylogeny (k=5) and whole genome phylogeny (k=4) for the original 13 phages (Fig. 2a-c).

The grouping results from the three classification methods are summarized in Fig. 2d-f. While each method produced somewhat different groupings, there were notable consistencies. For example, ΦWec172, ΦWec174, and ΦWec177 were consistently grouped together across all three methods (P2, T1, W1). Similarly, ΦWec189, ΦWec191, ΦWec193, and ΦWec196 formed a consistent group (P1, T2, W2).

When including the three additional phages, the optimal cluster numbers increased to 5, 6, and 5 for physiological, tail fiber, and whole genome classifications, respectively.

The similarity between these groupings and the phage groups based on target receptors was evaluated using the Adjusted Rand Index (ARI), which measures the agreement between two clustering methods (Table 6). All three classification methods showed high correlation with the grouping based on target receptors, both before and after the addition of the three new phages. Notably, the addition of phages with distinct receptor specificity improved the ARI values for all three methods, demonstrating the robustness of these classification approaches.

The whole genome-based classification consistently showed the highest similarity with target receptor grouping (ARI = 0.90 for 13 phages, increasing to 0.94 for 16 phages). Physiological characteristic-based clustering showed substantial improvement with the addition of new phages (ARI increasing from 0.82 to 0.89), demonstrating its ability to accurately classify phages with diverse receptor specificities.

Furthermore, we evaluated the similarity between the different classification methods themselves (Table 7). All three methods showed high concordance with each other, with ARI values ranging from 0.80 to 0.95. The addition of the three phages further increased the agreement between methods, indicating that as the diversity of phages increases, these different classification approaches tend to converge in their grouping patterns.

These results demonstrate that all three classification methods reflect the underlying biological relationships between phages, particularly their target receptor specificity. Importantly, the physiological characteristic-based classification, which can be determined without extensive genomic analysis, provides a reliable basis for estimating both genetic relationships and receptor specificity of phages across a diverse range of phage types.

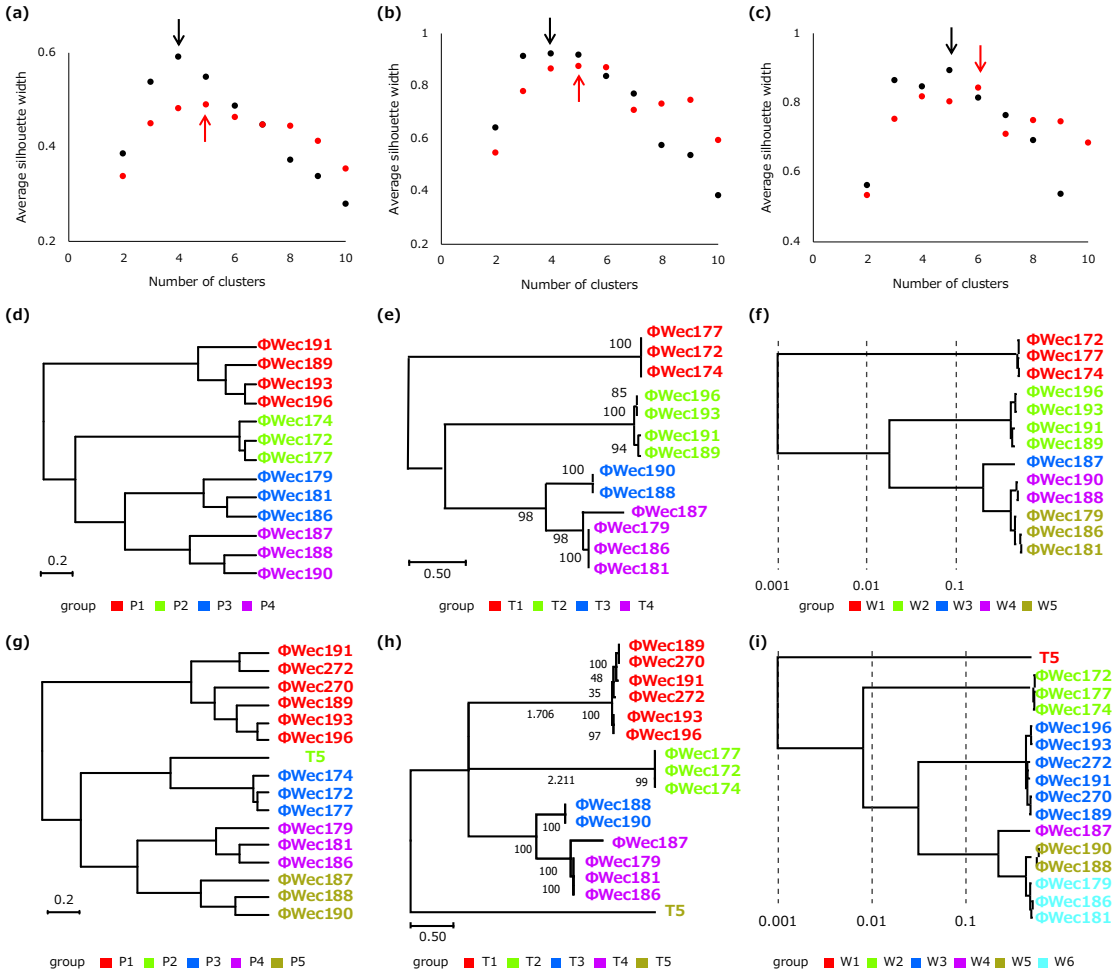


Fig. 2 Silhouette coefficient analysis and hierarchical clustering of phages by different methods

(a-c) Silhouette coefficient analysis for determining optimal cluster numbers based on (a) physiological characteristics, (b) tail fiber phylogeny, and (c) whole genome phylogeny. The plots show average silhouette width values for different numbers of clusters (k) from 2 to 10. Black plots represent analysis of 13 original phages, while red plots show results after addition of three supplementary phages (ΦWec270, ΦWec272, and T5). Arrows indicate the maximum silhouette coefficient values, representing the optimal number of clusters for each method. Note that in panel (c), results for k=10 are not shown as the coefficient value diverged to negative infinity for whole genome phylogeny. (d-f)

Hierarchical clustering dendrograms for the original 13 phages based on (d) physiological characteristics, (e) tail fiber phylogeny, and (f) whole genome phylogeny. (g-i) Hierarchical clustering dendrograms after addition of three supplementary phages based on (g) physiological characteristics, (h) tail fiber phylogeny, and (i) whole genome phylogeny. Colors in dendrograms (d-i) represent clusters determined using the optimal number identified by silhouette coefficient analysis in panels (a-c).

Table 6 ARI Values of Similarity between each Clustering Method with Target Receptor-Based Grouping Before and After Additional Phages

	ARI	
Comparison Pair	13 phages	16 phages
Physiological vs Target receptor	0.82	0.89
Tail fiber vs Target receptor	0.78	0.86
Whole genome vs Target receptor	0.90	0.94

Table 7 ARI Values of Similarity between Three Different Classification Methods Before and After Additional Phages

	ARI	
Comparison Pair	13 phages	16 phages
Physiological vs Tail fiber	0.80	0.87
Physiological vs Whole genome	0.91	0.95
Tail fiber vs Whole genome	0.87	0.92

429 **Discussion**

430 While phage therapy using a cocktail of phages targeting different receptors is known to
431 be effective, methods to identify phage target receptors remain limited and time-
432 consuming. In our previous study, we demonstrated that effective phage cocktails could
433 be prepared by grouping phages based on their physiological characteristics, but the actual
434 receptors of the phages were not confirmed (9). In this study, we validated our previous
435 clustering method through detailed analysis of each phage's target receptor. Our findings
436 demonstrate that phages with different target receptors can be grouped based on their
437 physiological characteristics, tail fiber sequences, or whole genomes. These results
438 provide a foundation for more efficient creation of diverse phage cocktails and deepen
439 our understanding of phage-host interactions, enabling rapid preparation of effective
440 therapeutic combinations.

441 **Phage target receptors**

442 In this study, we identified the target receptors of each phage in detail by combining the
443 analysis of phage-resistant strains with LPS structure analysis. These results reveal the
444 diversity and specificity of phage host recognition mechanisms and provide new insights
445 into the relationship between LPS structure and phage infectivity (Fig. 3).

446 The results of LPS structure analysis clearly show the specific effects of each gene
447 mutation on LPS structure. The complete disappearance of the upper band in DOC-PAGE
448 for *waaV* mutants (R172, R177-1, and R177-2) strongly suggests that in the TK001 strain,
449 the O-antigen is attached to the Glcβ added by WaaV. Additionally, the downward shift
450 of the R-core bands in *waaW* and *waaT* mutants (R174 and R181) indicates that the
451 deficiency of these genes results in a shorter R-core length. Interestingly, in the *waaG*
452 mutant (R190), the upper band almost disappeared, and the R-core band separated into

453 two. Since *waaG* is an early gene in outer core synthesis, this result suggests that while
454 R-core synthesis was incomplete, some compensatory mechanism may have led to the
455 expression of a unique LPS, consistent with previous observations by Klein et al. who
456 demonstrated that *waaG* deletion in *E. coli* K-12 resulted in increased incorporation of a
457 third Kdo residue and rhamnose, leading to accumulation of different structures of
458 glycoforms (16). On the other hand, no clear changes in band pattern were observed in
459 *waaY* mutants (R193 and R196), indicating that WaaY, which is involved only in Hep
460 phosphorylation, has little effect on the overall length of LPS but creates functionally
461 critical modifications for specific phage recognition. Notably, although ΦWec172, 174,
462 and 177 were suggested to be similar in our previous report (10) due to their common
463 isolation source, similar lysis range, and over 99% homology in ANI-based similarity
464 evaluation (Table S3), ΦWec172 did not require WaaY's function, unlike ΦWec174 and
465 177. These phages have two types of tail fibers: Tail fiber 1 consisting of 844 amino acid
466 residues and Tail fiber 2 consisting of 390 amino acid residues. There are the N777S
467 mutation in Tail fiber 1 of ΦWec174 and the K286E mutation in Tail fiber 2 of ΦWec177.
468 These differences suggest that ΦWec172 may interact with LPS differently from the other
469 two phages and may not require Heptose phosphorylation by WaaY. This observation
470 demonstrates the complexity and flexibility of phage receptor recognition and may
471 provide an example of adaptation strategies in phage-host interactions. However, as no
472 unique mutations were found in these tail fibers for ΦWec172 alone, further genomic
473 analysis, structural analysis, or mutation experiments are necessary to elucidate the
474 mechanism.
475 The changes in LPS structure were closely related to the infectivity of each phage. The
476 most important finding was that all phages were unable to infect the *waaG* mutant (R190),

477 clearly demonstrating that LPS is essential for infection by all phages used in this study.
478 Our structural prediction analyses revealed how different types of mutations in LPS
479 biosynthesis genes lead to diverse effects on receptor function. For extensive truncations
480 in WaaG, WaaV, WecA and ManB-1, the complete loss of catalytic domains explains the
481 absence or severe disruption of LPS structures observed in DOC-PAGE. In contrast, for
482 WaaW, WaaT, and WaaY mutations, the truncations appeared less extensive but
483 potentially significant to function. These mutations preserved most catalytic residues but
484 affected C-terminal regions containing clusters of basic amino acids. These positively
485 charged regions likely interact with the negatively charged LPS, suggesting how even
486 limited truncations could substantially alter substrate binding capabilities without
487 complete loss of enzymatic activity.
488 The WaaY mutations provided particularly compelling evidence for this mechanism.
489 DOC-PAGE analysis showed no visible changes in overall LPS length, suggesting the
490 core structure remained intact. However, phage infection experiments clearly
491 demonstrated functional changes, as certain phages (ΦWec189, ΦWec191, ΦWec193, and
492 ΦWec196) could not infect these mutants. Complementation with wild-type WaaY
493 restored phage sensitivity, confirming the direct relationship between the mutation and
494 phenotype. Our structural predictions explain this apparent discrepancy - WaaY's kinase
495 activity phosphorylates heptose residues without altering core length, creating negatively
496 charged regions that serve as critical recognition sites for specific phages. This
497 demonstrates the remarkable discriminatory capacity of phages, which can evolve to
498 recognize highly specific chemical modifications on the bacterial surface.
499 Furthermore, interesting results were obtained regarding the target receptor recognition
500 of ΦWec172, 174, and 177. These phages could not infect the *waaW* mutant but could

infect the strain with a mutation in WaaT, which functions just before WaaW in the LPS synthesis pathway. In the LPS synthesis pathway, WaaT transfers Galactose, followed by WaaW transferring second Galactose. This result suggests that these phages primarily target the Galactose transferred by WaaW. However, the fact that they can also infect the WaaT mutant suggests that similar structures, such as Glucose transferred by WaaO, may serve as alternative targets. This flexible recognition mechanism might be one of the phages' strategies for adapting to their hosts.

Our variable results in complementing *waaV* mutations highlight the complexity of LPS biosynthesis pathways. The successful complementation in R172 (deletion at position 451 bp) but not in R177-1 and R177-2 (deletions at position 525 bp) suggests that the position of the mutation within *waaV* may lead to different effects on downstream genes or LPS assembly mechanisms. The failure of mutant *waaV* variants to affect EoP when expressed in wild-type TK001 indicates that these mutations are likely recessive and do not interfere with wild-type function, pointing to potential secondary effects in the genetic background of specific mutants.

The addition of phages with known receptors (ΦWec270, ΦWec272, and T5) further validated our approach. ΦWec270 and ΦWec272, which share the same genus as ΦWec189, 191, 193, and 196, showed identical infection patterns, confirming they target the same receptor structures requiring both NfrB and phosphorylated heptose in LPS. In contrast, T5, which targets FhuA in *E. coli* K-12, remained infective against all our receptor mutant strains, indicating it recognizes a receptor not affected by our mutations. This diversity in receptor recognition mechanisms demonstrates the importance of using multiple phages targeting different host receptors in therapeutic applications.

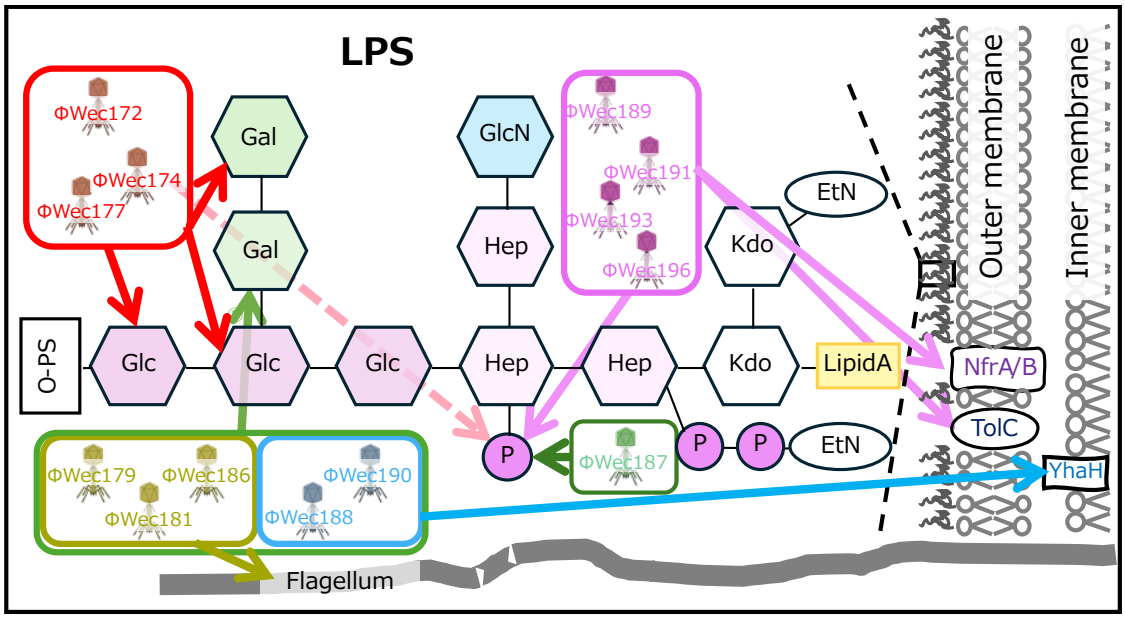


Fig. 3 Target receptors of each phage

The receptor diagram illustrates the diversity of receptors including LPS components, membrane proteins, and flagella. NfrA/B are depicted together as they represent functionally linked membrane proteins that cooperatively serve as phage receptors (18). Arrows extending from each phage indicate their target receptors. The dotted red arrow from ΦWec172 indicates that this phage does not require WaaY-mediated heptose phosphorylation for infection, despite being grouped with ΦWec174 and ΦWec177 based on genomic and physiological characteristics. This suggests subtle differences in receptor recognition mechanisms even among closely related phages.

Phage classification methods

In this study, we compared three approaches for phage classification: physiological characteristics, tail fiber phylogeny, and whole genome phylogeny. We enhanced our previously proposed physiological classification method through several methodological improvements, including optimized bacterial panel selection, implementation of Gower

distance for mixed data types to properly handle both qualitative infection data and quantitative measurements. Additionally, we applied silhouette coefficient analysis for objective determination of optimal cluster numbers across all classification methods, eliminating the subjective nature of cluster number selection that was a limitation in previous work.

Comparison of these methods with grouping based on target receptors showed that all methods performed better than random classification, with Adjusted Rand Index (ARI) values well above zero. All methods showed high agreement with target receptor-based classification, demonstrating strong interchangeability between these approaches. The improvement in ARI values after adding phages with distinct receptor specificities (ΦWec270, ΦWec272, and T5) further validates the robustness of our classification approaches. The proper clustering of T5 as a separate group across all methods demonstrates their ability to distinguish phages with unique receptor specificities, while the consistent grouping of ΦWec270 and ΦWec272 with phages sharing the same receptors confirms the methods' sensitivity to receptor-related characteristics.

While overall agreement between methods was high, some phages showed variable classification depending on the method used. Specifically, ΦWec187 was grouped differently across methods - clustering with ΦWec188 and ΦWec190 in physiological-based grouping, with ΦWec179, 181, and 186 in tail fiber analysis, and forming its own distinct group in whole genome classification. This variability persisted even after adding new phages to the analysis, suggesting that ΦWec187 possesses a unique combination of characteristics that are captured differently by each method. Such observations highlight that each classification approach may emphasize different aspects of phage biology.

Importantly, none of the methods completely matched the classification based on target

565 receptors. This suggests that phage target receptor specificity is a complex characteristic
566 that cannot be fully explained by a single classification criterion. The recognition of target
567 receptors is likely determined by not only specific regions of phage tail fibers but also
568 other structural proteins. Furthermore, subtle differences such as structural elements may
569 play significant roles that cannot be fully captured by whole genome analysis or
570 physiological characteristics alone.
571 For example, studies on bacteriophage T4 have revealed that structures such as the collar
572 and whiskers act as environment-sensing devices, regulating the retraction of long tail
573 fibers under unfavorable conditions and thus preventing infection, demonstrating that
574 phage-host recognition can involve structural elements beyond tail fibers to modulate
575 infection processes (17). These results highlight the challenges in phage classification and
576 functional prediction, and reflects the complexity of phage-host coevolution.
577 Nevertheless, this study demonstrated that our improved clustering based on
578 physiological characteristics is an effective method that reflects phage target receptor
579 specificity to a high degree. The high correlation between classifications based on
580 phenotypic data and genomic information supports the validity of our approach for
581 practical applications in phage selection, providing an important foundation for the
582 development of effective phage therapy.
583 **Limitations and future directions**
584 While our findings provide strong support for the validity of physiological characteristic-
585 based clustering as a proxy for target receptor diversity, our analyses were conducted
586 using a single *E. coli* strain (TK001) as the primary host, and the identified receptor
587 mutations may be specific to this genetic background. The high agreement between our
588 different classification methods suggests that physiological characteristics can serve as

589 reliable indicators of genetic relationships and receptor specificity. This has important
590 practical implications, particularly for phages targeting bacterial species with limited
591 genomic characterization. As noted in our introduction, while genomic approaches have
592 become increasingly accessible for well-studied organisms like *E. coli*, receptor
593 identification based on genomic data requires well-characterized bacterial genomes. For
594 emerging multidrug-resistant pathogens with less characterized genomes, our
595 physiological classification approach could provide a practical alternative that does not
596 depend on extensive genomic information.

597 In conclusion, this study provides important insights into *E. coli* phage target receptors
598 and the relationship between various classification methods. Our findings reveal that
599 phage host recognition mechanisms are diverse and specific, with many phages
600 recognizing specific structures of LPS, while others target membrane proteins or flagella.
601 Through robust molecular analysis of 13 phages, including detailed characterization of
602 their physiological features, whole genome sequences, and target receptors, we have
603 definitively validated our previously proposed classification method by demonstrating its
604 correlation with actual phage target receptors.

605 While our findings are based on *E. coli* phages, this approach has potential broader
606 applications. Further studies with phages infecting other bacterial species will be
607 necessary to confirm the universal applicability of these methods. Nevertheless, in
608 clinical settings where conventional receptor identification is impractical and rapid
609 decision-making is crucial, our validated classification approach offers powerful tools for
610 creating effective phage cocktails that can suppress the emergence of phage-resistant
611 bacteria, marking a significant advancement in the field of phage therapy.

612

613 **Materials and methods**

614 **Reagents and bacteria**

615 Luria–Bertani (LB) medium (BD Difco, Franklin Lakes, NJ, USA) and the soft agar
616 containing 0.5% agar were used for bacterial and phage cultivation. *E. coli* strain TK001,
617 originally isolated from dextran sodium sulfate (MP Biomedicals, Soho, OH, USA)-
618 induced colitis mouse feces, was used as the host for phage screening in this study (10).

619 **Phage preparation**

620 Thirteen phages (ΦWec172, ΦWec174, ΦWec177, ΦWec179, ΦWec181, ΦWec186,
621 ΦWec187, ΦWec188, ΦWec190, ΦWec189, ΦWec191, ΦWec193, and ΦWec196) were
622 propagated using *E. coli* TK001 as the host through the plate lysate method.

623 For validation experiments, three additional phages were included. ΦWec270 and
624 ΦWec272, previously characterized in our taxonomic study (15), and bacteriophage T5
625 (NBRC 20005) from the Biological Resource Center, National Institute of Technology
626 and Evaluation (NBRC), were used. These three phages can infect TK001, but were
627 propagated using *E. coli* MG1655 (CGSC #6300) as the host using the same plate lysate
628 method.

629 Briefly, 100 μL of overnight-cultured TK001 and 100 μL of phage suspension (10^4
630 plaque-forming unit (PFU)/mL) were added to soft agar (0.5% agar) and overlaid on agar
631 plates. After overnight incubation at 37 °C, 5 mL of SM buffer (0.1 M NaCl, 8 mM
632 $MgSO_4 \cdot 7H_2O$, 50 mM Tris-HCl buffer [pH=7.5], 0.1% gelatin) was added to each plate
633 and subjected to orbital shaking (approximately 250 rpm, 2-3 hours). The supernatant was
634 centrifuged (9,000 rpm, 15 min, 4 °C) to separate solid and liquid components. The
635 supernatant was transferred to a new tube, and 1 mL of chloroform was added. After
636 vortexing and settling, and purified phage was decanted and served as the phage stock.

637 **Generation and selection of phage resistant receptor mutants**

638 Phage-resistant strains were generated using both liquid and solid culture methods:

639 *Liquid culture method*: Log-phase TK001 culture was diluted 100-fold, and 100 μL of
640 high-titer phage suspension (>10^8 PFU/mL, MOI > 10) was added. The mixture was
641 incubated at 37°C with shaking at 40 rpm in a compact rocking incubator (TVS062CA,
642 Advantec Toyo Co. Ltd., Tokyo, Japan). The emergence of resistant bacteria was
643 confirmed by turbidity curves the following day. The culture was centrifuged (9000 rpm,
644 5 min, 4 °C), and the pellet was washed with 1× phosphate-buffered saline (PBS). The
645 pellet was serially diluted in 1× PBS (1 mL), and 20 μL of each dilution was spotted onto
646 soft agar containing 100 μL of high-titer phage. Eight phage-resistant *E. coli* colonies
647 were isolated for each phage and purified by streak culture at least twice. Purified colonies
648 were inoculated into 1 mL of LB, cultured for 11 hours, and stored at -80 °C.

649 *Solid culture method*: 100 μL of overnight-cultured TK001 and 100 μL of high-titer phage
650 suspension (>10^8 PFU/mL) were added to soft agar and overlaid on agar plates. Four
651 colonies per phage (one for ΦWec190 and three for ΦWec193) were isolated the
652 following day and purified by streak culture at least twice. Purified colonies were
653 processed and stored as described in the liquid culture method.

654 Phage adsorption assay was used to select target receptor mutants from the acquired
655 phage-resistant strains. A mixture of overnight-cultured resistant strain (100 μL) and LB
656 (890 μL) was combined with phage suspension (>10^7 PFU/mL, 100 μL). Samples (10 μL)
657 were taken at 0- and 10-min post-phage addition, mixed with 990 μL SM buffer (pre-
658 added with 100 μL chloroform), and centrifuged (15,000 rpm, >2 min, 4 °C). The
659 supernatant (100 μL) and overnight-cultured TK001 (100 μL) were mixed in soft agar
660 and overlaid on agar plates. After overnight incubation, the unadsorbed ratio was

661 calculated by dividing the number of plaques at 10 min by the number at 0 min. Strains
662 with an adsorption rate (100 - unadsorbed ratio) below 20% were designated as target
663 receptor mutants.

664 **Whole genome analysis of wild type strain**

665 DNA was extracted from overnight static cultures of *E. coli* TK001 in LB medium using
666 ISOPLANT II (Nippon Gene, 310-04151, Tokyo, Japan) according to the manufacturer's
667 instructions.

668 Sequencing was performed by BGI Genomics. Continuous genome sequencing was
669 assembled using two different sequencing technologies. Shotgun sequencing was
670 performed using DNBSEQ G400 (MGI) with a read length of 150 bp. Clean data was
671 generated using SOAPnuke (BGI) to remove adapters and low-quality reads (parameters:
672 "-n 0.01 -l 20 -q 0.4 --adaMis 3 --outQualSys 1")(19). Long reads obtained by Oxford
673 Nanopore Technologies were processed using porechop (ver. 0.2.4, default parameters).
674 The processed shotgun and long-read data were assembled using SPAdes (ver. 3.15.5,
675 parameters: "-k auto --cov-cutoff auto --careful -t 12") for hybrid assembly (20).
676 Annotation of the resulting contigs was performed using prokka (21). Basic statistical
677 analysis of the contigs was conducted using seqkit (22).

678 **Whole genome analysis and mutation detection in mutant strains**

679 DNA was extracted from selected target receptor mutant strains as described for the wild-
680 type strain. BGI performed shotgun sequencing and read processing similarly to the wild-
681 type strain. The mutant strain shotgun reads were mapped to the indexed wild-type
682 sequence using bwa. Samtools was used to convert the mapping results to BAM format
683 and sort them. The sorted BAM files were indexed using samtools. Guided assembly was
684 performed using pilon (ver. 1.24, parameters: "--vcf --tracks --changes --verbose") with

685 the index files (23). Non-synonymous mutations were detected by comparing each
686 resistant strain's sequence obtained through guided assembly with the wild-type strain
687 using nucdiff (24).

688 **Complementation experiments**

689 To confirm that the identified mutations in phage-resistant strains were directly
690 responsible for the resistance phenotypes, we performed complementation experiments.
691 Wild-type genes corresponding to the mutated genes in resistant strains were cloned into
692 the low-copy plasmid pACYC184. This plasmid was selected for its stability in TK001,
693 and dual antibiotic resistance markers (tetracycline and chloramphenicol).

694 Each gene of interest, along with its native promoter region, was PCR-amplified from
695 TK001 genomic DNA. The amplified fragments were designed to replace the tetracycline
696 resistance gene in pACYC184 (Nippon Gene Co., Ltd., Tokyo, Japan) using In-Fusion
697 cloning (Takara Bio Inc., Shiga, Japan). Care was taken to avoid including unnecessary
698 adjacent genes when incorporating promoter regions. The resulting constructs were
699 initially transformed into *E. coli* JM109 (Takara Bio Inc., Shiga, Japan), and positive
700 transformants were selected on LB agar with chloramphenicol (25 μg/mL). For each gene,
701 approximately 10 colonies were isolated and verified by PCR amplification of the inserted
702 gene and by confirming loss of tetracycline resistance.

703 Verified plasmid constructs were then isolated and transformed into the corresponding
704 phage-resistant mutant strains using a standard chemical competent cell protocol (see also
705 supplementary method). For each complementation, three independent transformants
706 were isolated and verified by PCR to confirm the presence of the complementing gene.
707 The complemented strains were then tested for restored phage sensitivity using spot tests
708 as described in the "Spot test using target receptor mutants" section. EoP values were

709 calculated relative to the wild-type strain TK001 carrying the empty pACYC184 vector.
710 Biological replicates were performed 3 times.
711 The complemented strains were then tested for restored phage sensitivity using the same
712 spot test method as described in the "Spot test using target receptor mutants" section. EoP
713 values were calculated relative to the wild-type strain TK001 carrying the empty
714 pACYC184 vector. Biological replicates were performed 3 times.
715 For *waaV* mutations, which showed variable complementation results, we performed
716 additional experiments by expressing the mutant variants of *waaV* in wild-type TK001 to
717 assess whether the mutant proteins might interfere with wild-type function.
718 **Protein structure prediction and analysis**
719 To understand the molecular basis of how the identified mutations affect protein function,
720 we predicted protein structure using AlphaFold2 via Google Colaboratory (25)(26) or
721 AlphaFold3 (27). For each mutated protein (WaaG, WaaV, WaaW, WaaT, WaaY, WecA,
722 and ManB-1), we generated structural models of wild-type. The predictions were run with
723 default parameters, and models were evaluated using the predicted local distance
724 difference test (pLDDT) and predicted aligned error (pAE).
725 For WaaG, where crystal structures of homologs are available, predicted WaaG structure
726 was superimposed on a previously solved crystal structure (PDB: 2IW1) to identify
727 catalytic and substrate-binding sites. For WaaY and ManB1, the complex with ATP or
728 GTP were predicted using AlphaFold3. For other proteins, potential binding sites were
729 predicted using AutoDock4.2.6
730 (28) with UDP-glucose or UDP-galactose as ligands, as appropriate for each enzyme's
731 function.
732 Electrostatic surface potentials were calculated using the Adaptive Poisson-Boltzmann

733 Solver (APBS) program.

734 All structural analyses and visualizations were performed using PyMOL (version 3.1.4.1).

735 **Preparation of dried defatted bacterial cells**

736 Dried defatted bacterial cells were prepared for LPS extraction. Bacterial cells were
737 cultured in 2 L of medium and harvested by centrifugation (4,800 rpm, 15 min, 4°C) in
738 four 500 mL centrifuge tubes. From *E. coli* cultures, approximately 20 g of wet cells were
739 typically obtained. Cell pellets were washed by resuspension in 30 mL of 1× PBS and
740 collected by centrifugation (4,800 rpm, 15 min, 4°C) in 50 mL tubes.

741 The cells were then defatted through sequential washing steps. First, cells were
742 resuspended in 30 mL of 100% ethanol using a spatula, followed by centrifugation (4,800
743 rpm, 15 min, 4°C). This ethanol washing step was repeated until the cell pellet lost its
744 sticky consistency (typically three times), with the supernatant turning yellow. The pellet
745 was then washed once with 30 mL acetone, using gentle deceleration during
746 centrifugation to prevent pellet disruption. Finally, cells were washed with 30 mL of
747 hexane or diethyl ether.

748 The resulting pellet was dried under vacuum in a desiccator for 1-3 days. This process
749 typically yielded 4-5 g of off-white to white dried defatted bacterial cells from *E. coli*
750 cultures.

751 **LPS structure analysis**

752 LPS structure analysis of wild type and target receptor mutant strains was performed
753 using two methods: pathway analysis to predict the LPS synthesis pathway from retained
754 genes, and sodium deoxycholate-polyacrylamide gel electrophoresis (DOC-PAGE) for
755 LPS structure analysis.

756 *Pathway analysis*: KEGG Mapper was used to map retained genes to known synthesis

757 pathways based on Kegg Orthology (KO) assignment results obtained using
758 BlastKOALA (29)·(30)·(31).
759 *Structural analysis*: To obtain LPS from bacterial cells, dried defatted bacterial cells were
760 first prepared. LPS extraction from dried defatted bacterial cells was performed using the
761 hot phenol-water extraction method (32). Extracted LPS was visualized by DOC-PAGE
762 followed by silver staining (33). Silver staining was performed using the Silver Stain MS
763 Kit (Fujifilm Wako, 299-58901, Osaka, Japan), with a modified sodium periodate
764 oxidation step between first and second fixative solutions (34). The gel was shaken in 10
765 times its volume of 1% w/v sodium periodate solution for 20 min, followed by three 5-
766 min washes with MilliQ water. As control samples for the LPS structure of TK001, LPS
767 extracted from the following strains was used: *Salmonella enterica* serovar Typhimurium
768 (Sigma; catalogue no. L6511), *Escherichia coli* O111:B4 (Sigma; catalogue no. L4130),
769 and various *S. enterica* serovars with different chemotypes (LPSs were extracted for
770 bacteria of kindly provided by Dr. Hajime Ikigai). For presentation in Fig. 1(c), lanes
771 from the original gel were rearranged for logical comparison without altering the relative
772 positions or intensities of the bands. The original, unedited gel images were included in
773 the supplementary materials (Fig. S3) for reference, ensuring transparency in the data
774 presentation.
775 **Spot test using target receptor mutants**
776 Each phage was serially diluted 10-fold up to 10^{-7} and spotted (1 μL) onto target
777 receptor mutant lawns (35). The efficiency of plating (EoP) of the mutants relative to the
778 wild type was calculated. Biological replicates were performed 3-7 times.
779 **Spot test using the Keio collection**
780 For the five *E. coli* phages (ΦWec187, 189, 191, 193, and 196) capable of infecting K-12

781 strains(10), spot tests were conducted using the Keio collection (NBRP, Japan), a
782 comprehensive gene knockout library (36)·(37). The EoP of Keio collection strains
783 relative to the wild-type BW25113 was calculated. The Keio collection strains used were
784 limited to those predicted to be related to the target receptors of ΦWec187, 189, 191, 193,
785 and 196 based on phage-resistant TK001 (Table S1). Biological replicates were
786 performed 3 times.

787 **Whole genome construction of phages**

788 Sequencing data were assembled using Velvet software (ver. 1.2.10) (38).

789 **Phylogenetic tree based on tail fibers**

790 Whole genome information of each phage was annotated using RAST, and tail fiber
791 sequences were extracted (39). Additionally, tail fibers of the closest relatives were
792 obtained by BLAST search of each phage's whole genome, and similar sequences were
793 extracted from the whole genomes of phages drawing these sequences. The obtained
794 sequences were translated into amino acids, and for each phage, the sequences were
795 artificially combined into a single sequence. A phylogenetic tree was constructed using
796 MEGA X with the Maximum-likelihood method and 1,000 bootstrap replicates (40).

797 **Whole genome phylogenetic tree and similarity**

798 The whole genome phylogenetic tree was created using Viptree (41). Furthermore,
799 phylogenetic analysis of the phages was conducted using vConTACT2 (42) with the
800 reference genome of 3503 phage (ProkaryoticViralRefSeq201) from vConTACT2, and
801 visualized using Cytospace v3.10.1 (43). The Average Nucleotide Identity (ANI) was
802 calculated using pyani (44). The parameters used were: pyani: anib and ANI calculator:
803 minimum identity 0%, with other parameters set to default.

804 **Phage grouping based on physiological characteristics, tail fibers, and whole**

805 **genomes**

806 Phages were grouped based on their physiological characteristics as previously reported.
807 To improve the robustness of our previously proposed clustering method, we optimized
808 the bacterial panel for host range evaluation by analyzing infection patterns of 58 phages
809 (including phages previously reported in our previous study (15)) against 41
810 *Enterobacteriaceae* strains (Table S5). Using a greedy algorithm, we selected 16 strains
811 that collectively maximized the discriminatory power while minimizing the number of
812 strains required.

813 For data analysis, Gower distance was implemented instead of Euclidean distance to
814 appropriately handle the mixed nature of our dataset, which contained both qualitative
815 infection data (binary host range patterns) and quantitative measurements (continuous
816 physiological parameters).

817 To objectively determine the optimal number of clusters, silhouette coefficient analysis
818 was performed for all three grouping methods (physiological characteristics, tail fiber
819 phylogenetics, and whole genome phylogenetics). For each potential number of clusters
820 k (ranging from 2 to 9), we calculated the average silhouette coefficient across all phages
821 in the dataset. The silhouette coefficient for each phage measures how similar it is to its
822 own cluster compared to other clusters, with values ranging from -1 to 1. Higher values
823 indicate better-defined clusters. The optimal number of clusters for each grouping method
824 was determined as the value of k that maximized the average silhouette coefficient.

825 These analyses were performed using R (version 4.0.3) with the "cluster" and "factoextra"
826 packages for silhouette calculations. Hierarchical clustering was performed using the
827 'complete' linkage method as previously reported (10).

828 **Comparison of similarity between methods**

Data analysis was performed using Python 3.9.12, with major packages including pandas (1.4.2), numpy (1.22.3), and scikit-learn (1.0.2). The similarity between different grouping methods and with target receptor-based grouping was evaluated using Adjusted Rand Index (ARI). Label normalization was performed in the similarity evaluation to improve the accuracy of comparison between methods. These analyses were executed on Linux.

Acknowledgements

We extend our sincere gratitude to Midori Ishikawa and Hirohito Wakashima for providing the LPS samples used as controls in our DOC-PAGE analysis. Their contributions were crucial for our comparative study of *E. coli* LPS structures. Specifically, we acknowledge:

Midori Ishikawa for isolating LPS from *S. enterica* serovar Typhimurium TV119 (Ra chemotype) and *S. enterica* serovar Typhimurium SL733 (Rb_1 chemotype).

Hirohito Wakashima for isolating LPS from *S. enterica* serovar Typhimurium SL1034 (semi-rough chemotype), *S. enterica* serovar Minnesota SF1119 (RcP^- chemotype), *S. enterica* serovar Minnesota SF1121 (Rd_1P^- chemotype), *S. enterica* serovar Minnesota SF1118 (Rd_2 chemotype), and *S. enterica* serovar Typhimurium TA2168 (Re chemotype).

In addition to the previous acknowledgement of Matthew Imanaka for English language editing on the initial version, we would like to thank Claude 3.7 Sonnet (Anthropic) for its assistance in improving the English grammar and style of this revised manuscript.

FUNDING INFORMATION

This research was supported by a Waseda University Grant for Special Research Projects

(Project number: 2022C-160).

References

1. Chen P, Sun H, Ren H, Liu W, Li G, Zhang C. 2020. LamB, OmpC, and the Core Lipopolysaccharide of *Escherichia coli* K-12 Function as Receptors of Bacteriophage Bp7. J Virol 94:1–14.
2. Tanji Y, Shimada T, Yoichi M, Miyanaga K, Hori K, Unno H. 2004. Toward rational control of *Escherichia coli* O157:H7 by a phage cocktail. Appl Microbiol Biotechnol 64:270–274.
3. Yang Y, Shen W, Zhong Q, Chen Q, He X, Baker JL, Xiong K, Jin X, Wang J, Hu F, Le S. 2020. Development of a Bacteriophage Cocktail to Constrain the Emergence of Phage-Resistant *Pseudomonas aeruginosa*. Front Microbiol 11:327.
4. Tanji Y, Shimada T, Fukudomi H, Miyanaga K, Nakai Y, Unno H. 2005. Therapeutic use of phage cocktail for controlling *Escherichia coli* O157:H7 in gastrointestinal tract of mice. J Biosci Bioeng 100:280–287.
5. Bai J, Jeon B, Ryu S. 2019. Effective inhibition of *Salmonella* Typhimurium in fresh produce by a phage cocktail targeting multiple host receptors. Food Microbiol 77:52–60.
6. Schooley RT, Biswas B, Gill JJ, Hernandez-Morales A, Lancaster J, Lessor L, Barr JJ, Reed SL, Rohwer F, Benler S, Segall AM, Taplitz R, Smith DM, Kerr K, Kumaraswamy M, Nizet V, Lin L, McCauley MD, Strathdee SA, Benson CA, Pope RK, Leroux BM, Picel AC, Mateczun AJ, Cilwa KE, Regeimbal JM, Estrella LA, Wolfe DM, Henry MS, Quinones J, Salka S, Bishop-Lilly KA, Young R, Hamilton T. 2017. Development and use of personalized bacteriophage-based

877 therapeutic cocktails to treat a patient with a disseminated resistant *Acinetobacter*
878 *baumannii* infection. Antimicrob Agents Chemother 61:e00954-17.
879 7. Nakamura K, Fujiki J, Nakamura T, Furusawa T, Gondaira S, Usui M, Higuchi H,
880 Tamura Y, Iwano H. 2021. Fluctuating Bacteriophage-induced galU Deficiency
881 Region is Involved in Trade-off Effects on the Phage and Fluoroquinolone
882 Sensitivity in *Pseudomonas aeruginosa*. Virus Res 306:198596.
883 8. Liang J, Li X, Zha T, Chen Y, Hao H, Liu C, Duan R, Xiao Y, Su M, Wang X,
884 Jing H. 2016. DTDP-rhamnosyl transferase RfbF, is a newfound receptor-related
885 regulatory protein for phage phiYe-F10 specific for *Yersinia enterocolitica*
886 serotype O:3. Sci Rep 6:1–11.
887 9. Gröschel MI, Meehan CJ, Barilar I, Diricks M, Gonzaga A, Steglich M, Conchillo-
888 Solé O, Scherer IC, Mamat U, Luz CF, De Bruyne K, Utpatel C, Yero D, Gibert I,
889 Daura X, Kampmeier S, Rahman NA, Kresken M, van der Werf TS, Alio I, Streit
890 WR, Zhou K, Schwartz T, Rossen JWA, Farhat MR, Schaible UE, Nübel U, Rupp
891 J, Steinmann J, Niemann S, Kohl TA. 2020. The phylogenetic landscape and
892 nosocomial spread of the multidrug-resistant opportunist *Stenotrophomonas*
893 *maltophilia*. Nat Commun 11.
894 10. Kaneko T, Osaka T, Tsuneda S. 2023. Tailoring Effective Phage Cocktails for
895 Long-Term Lysis of *Escherichia coli* Based on Physiological Properties of
896 Constituent Phages. PHAGE 4:128–135.
897 11. Kaneko T, Uchiyama J, Osaka T, Tsuneda S. 2025. Novel *Escherichia coli* Phages
898 Representing a Distinct Genus within *Stephanstirmvirinae*: Genome and Host
899 Range Characteristics. bioRxiv 2025.02.05.636633.
900 12. K. H, V. B. 1978. Functional Interaction of the tonA/tonB Receptor System in

901 *Escherichia coli*. J Bacteriol 135:190–197.
902 13. Volkmar B. 2009. FhuA (TonA), the Career of a Protein. J Bacteriol 191:3431–
903 3436.
904 14. Vinogradov E V., Van Der Drift K, Thomas-Oates JE, Meshkov S, Brade H, Holst
905 O. 1999. The structures of the carbohydrate backbones of the lipopolysaccharides
906 from *Escherichia coli* rough mutants F470 (R1 core type) and F576 (R2 core type).
907 Eur J Biochem 261:629–639.
908 15. Kaneko T, Aiba Y, Miyanaga K, Cui L, Tsuneda S. 2025. A Phage Collection
909 Propagated in *Escherichia coli* K-12 Shows Broad Lytic Activity Against a
910 Diverse Panel of Multidrug-Resistant Isolates. PHAGE
911 https://doi.org/10.1089/phage.2024.0023.
912 16. Klein G, Lindner B, Brade H, Raina S. 2011. Molecular basis of
913 lipopolysaccharide heterogeneity in Escherichia coli: Envelope stress-responsive
914 regulators control the incorporation of glycoforms with a third 3-deoxy-α-D-
915 manno-oct-2-ulosonic acid and rhamnose. J Biol Chem 286:42787–42807.
916 17. Fokine A, Zhang Z, Kanamaru S, Bowman, D. V, Aksyuk, A. A, Arisaka F, Rao,
917 B. V, Rossmann, G. M. 2014. The Molecular Architecture of the Bacteriophage
918 T4 Neck. J Mol Biol 23:1–7.
919 18. H. JE, Regine H. 2021. A Novel Locally c-di-GMP-Controlled Exopolysaccharide
920 Synthase Required for Bacteriophage N4 Infection of Escherichia coli. MBio
921 12:e03249-21.
922 19. Chen Y, Chen Y, Shi C, Huang Z, Zhang Y, Li S, Li Y, Ye J, Yu C, Li Z, Zhang
923 X, Wang J, Yang H, Fang L, Chen Q. 2018. SOAPnuke: A MapReduce
924 acceleration-supported software for integrated quality control and preprocessing of

925 high-throughput sequencing data. Gigascience 7:1–6.
926 20. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin
927 VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin A V., Sirotkin A V., Vyahhi
928 N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: A new genome assembly
929 algorithm and its applications to single-cell sequencing. J Comput Biol 19:455–
930 477.
931 21. Seemann T. 2014. Prokka: Rapid prokaryotic genome annotation. Bioinformatics
932 30:2068–2069.
933 22. Shen W, Le S, Li Y, Hu F. 2016. SeqKit: A cross-platform and ultrafast toolkit for
934 FASTA/Q file manipulation. PLoS One 11:1–10.
935 23. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA,
936 Zeng Q, Wortman J, Young SK, Earl AM. 2014. Pilon: An integrated tool for
937 comprehensive microbial variant detection and genome assembly improvement.
938 PLoS One 9.
939 24. Khelik K, Lagesen K, Sandve GK, Rognes T, Nederbragt AJ. 2017. NucDiff: In-
940 depth characterization and annotation of differences between two sets of DNA
941 sequences. BMC Bioinformatics 18:1–14.
942 25. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O,
943 Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, Bridgland A, Meyer C, Kohl
944 SAA, Ballard AJ, Cowie A, Romera-Paredes B, Nikolov S, Jain R, Adler J, Back
945 T, Petersen S, Reiman D, Clancy E, Zielinski M, Steinegger M, Pacholska M,
946 Berghammer T, Bodenstein S, Silver D, Vinyals O, Senior AW, Kavukcuoglu K,
947 Kohli P, Hassabis D. 2021. Highly accurate protein structure prediction with
948 AlphaFold. Nature 596:583–589.

949 26. Mirdita M, Schütze K, Moriwaki Y, Heo L, Ovchinnikov S, Steinegger M. 2022.
950 ColabFold: making protein folding accessible to all. Nat Methods 19:679–682.
951 27. Abramson J, Adler J, Dunger J, Evans R, Green T, Pritzel A, Ronneberger O,
952 Willmore L, Ballard AJ, Bambrick J, Bodenstein SW, Evans DA, Hung C-C,
953 O'Neill M, Reiman D, Tunyasuvunakool K, Wu Z, Žemgulytė A, Arvaniti E,
954 Beattie C, Bertolli O, Bridgland A, Cherepanov A, Congreve M, Cowen-Rivers AI,
955 Cowie A, Figurnov M, Fuchs FB, Gladman H, Jain R, Khan YA, Low CMR, Perlin
956 K, Potapenko A, Savy P, Singh S, Stecula A, Thillaisundaram A, Tong C, Yakneen
957 S, Zhong ED, Zielinski M, Žídek A, Bapst V, Kohli P, Jaderberg M, Hassabis D,
958 Jumper JM. 2024. Accurate structure prediction of biomolecular interactions with
959 AlphaFold 3. Nature 630:493–500.
960 28. Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, Olson
961 AJ. 2009. AutoDock4 and AutoDockTools4: Automated docking with selective
962 receptor flexibility. J Comput Chem 30:2785–2791.
963 29. Kanehisa M, Sato Y, Morishima K. 2016. BlastKOALA and GhostKOALA:
964 KEGG Tools for Functional Characterization of Genome and Metagenome
965 Sequences. J Mol Biol 428:726–731.
966 30. Kanehisa M, Sato Y. 2020. KEGG Mapper for inferring cellular functions from
967 protein sequences. Protein Sci 29:28–35.
968 31. Kanehisa M, Sato Y, Kawashima M. 2022. KEGG mapping tools for uncovering
969 hidden features in biological data. Protein Sci 31:47–53.
970 32. Westphal O, Jann K, Himmelspach K. 1983. Chemistry and immunochemistry of
971 bacterial lipopolysaccharides as cell wall antigens and endotoxins. Prog Allergy
972 33:9–39.

973 33. Kokubo T, Galanos C. 1988. Analysis of Salmonella lipopolysaccharides by
974 sodium deoxycholate-polyacrylamide gel electrophoresis. J Chromatogr 450:381–
975 387.
976 34. Fomsgaard A, Freudenberg MA, Galanos C. 1990. Modification of the silver
977 staining technique to detect lipopolysaccharide in polyacrylamide gels. J Clin
978 Microbiol 28:2627–2631.
979 35. Mattila S, Ruotsalainen P, Jalasvuori M. 2015. On-Demand Isolation of
980 Bacteriophages Against Drug-Resistant Bacteria for Personalized Phage Therapy
981 Bacteria Strains and Culturing. Front Microbiol 6:1–7.
982 36. Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, Datsenko KA, Tomita
983 M, Wanner BL, Mori H. 2006. Construction of *Escherichia coli* K-12 in-frame,
984 single-gene knockout mutants : the Keio collection. Mol Syst Biol 2006.0008.
985 37. Yamamoto N, Nakahigashi K, Nakamichi T, Yoshino M, Takai Y, Touda Y,
986 Furubayashi A, Kinjyo S, Dose1 H, Hasegawa M, Datsenko KA, Nakayashiki T,
987 Tomita M, Wanner BL, Mor H. 2009. Update on the Keio collection of *Escherichia*
988 *coli* single-gene deletion mutants. Mol Syst Biol 5.
989 38. Zerbino DR, Birney E. 2008. Velvet: Algorithms for de novo short read assembly
990 using de Bruijn graphs. Genome Res 18:821–829.
991 39. Aziz RK, Bartels D, Best A, DeJongh M, Disz T, Edwards RA, Formsma K,
992 Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL,
993 Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich
994 C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O. 2008. The RAST
995 Server: Rapid annotations using subsystems technology. BMC Genomics 9:1–15.
996 40. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. 2018. MEGA X: Molecular

997 evolutionary genetics analysis across computing platforms. Mol Biol Evol
998 35:1547–1549.
999 41. Nishimura Y, Yoshida T, Kuronishi M, Uehara H, Ogata H, Goto S. 2017.
1000 ViPTree: The viral proteomic tree server. Bioinformatics 33:2379–2380.
1001 42. Bolduc B, Jang H Bin, Doulcier G, You Z, Roux S, Sullivan MB. 2017.
1002 vConTACT : an iVirus tool to classify double-stranded DNA viruses that infect
1003 Archaea and Bacteria 1–26.
1004 43. Paul S, Andrew M, Owen O, Baliga NS, Wang JT, Ramage D, Amin N,
1005 Schwikowski B, Ideker T. 1971. Cytoscape: A Software Environment for
1006 Integrated Models. Genome Res 13:426.
1007 44. Pritchard L, Glover RH, Humphris S, Elphinstone JG, Toth IK. 2016. Genomics
1008 and taxonomy in diagnostics for food security: Soft-rotting enterobacterial plant
1009 pathogens. Anal Methods 8:12–24.
1010