Title

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

Running title

Phenotype, genomic, and receptor analysis of coliphages

Keywords

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

Author's names

Tomoyoshi Kaneko[1,2], Toshifumi Osaka[3], Minoru Inagaki[4], Satoshi Tsuneda[1,2†]

Author's affiliations

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

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

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

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

Corresponding author

[†] 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

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 combined genomic analysis and lipopolysaccharide (LPS) structural analysis of phage-resistant *E. coli* strains to identify target receptors of the phages. 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. Notably, 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 into five identical groups. Meanwhile, classification based on target receptors placed them into six groups, which was very similar to the grouping pattern based on the three independent parameters described above. Our results suggest that phage phenotypes and genotypes are closely related to target receptor specificity and the physiological characteristics of *E. coli* phages. 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 initial phage screening and selection was completed within 18 hours to address the patient's critical condition, though additional phages were subsequently identified and administered to the patient. In such urgent cases, conventional methods based on receptor identification is impractical, and more rapid treatment decision-making is required[6].

Traditionally, phage target receptor identification has involved a labor-intensive process of creating phage-resistant bacteria, conducting whole-genome comparative analysis, and identifying mutation sites[7,8]. While this methodology is reliable, it is unsuitable for clinical application due to its time and cost requirements. As such, in personalized phage therapy, phages are often combined in therapeutic cocktails without the knowledge of each phage's receptor. If multiple phages target the same receptor, a single receptor mutation could render the cocktail ineffective.

In our previous study, we proposed a clustering method based on physiological characteristics such as growth dynamics and host range. This method makes it possible to quickly create effective phage cocktails without identifying the target receptors[9]. However, at the time, we did not evaluate the validity of the proposed clustering methodology.

The present study aimed to validate the proposed clustering method based on physiological characteristics through detailed analysis of each phage's target receptor. Furthermore, we compared this 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 previously proposed physiological clustering approach. These findings are expected to contribute to rapid treatment decision-making in clinical settings and the standardization of phage therapy.

Results

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

Hybrid assembly resulted in two contigs for the whole genome of the *E. coli* TK001 wild-type strain isolated from mouse feces. The total genome length was 4,779,140 bp with a GC content of 50.79%. One of the two contigs was 145 bp long and was manually deleted as it was deemed unnecessary for subsequent analyses.

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

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

phage used to make strains resistant	strain	variant part	description	mutation site in WT	mutation type
ΦWec172	R172	waaV	beta1,3-glucosyltransferase	3934566	Frame shift (A→.)
ΦWec174	R174	waaW	UDP-galactose--(galactosyl) LPS alpha1,2-galactosyltransferase	3935258	Frame shift (.→T)
ΦWec177	R177-1	waaV	beta1,3-glucosyltransferase	3934492	Frame shift (A→.)
ΦWec177	R177-2	waaV	beta1,3-glucosyltransferase	3934492	Frame shift (A→.)
		wecA	Undecaprenyl-phosphate alpha-*N*-acetylglucosaminyl 1-phosphate transferase	4108976	Frame shift (A→.)
ΦWec181	R181	waaT	LPS 1,2-glucosyltransferase	2178550-2178553	frame shift (CCTC→.)
		manB-1	Mannose-1-phosphate guanylyltransferase	3937043	Stop codon (C→T)
ΦWec186	R186	fihD	Flagellar transcriptional regulator FlhD	2021475	Stop codon (C→T)
		bcsG	Cellulose biosynthesis protein BcsG	3831107	Ser→Pro (A→G)
ΦWec187	R187	bcsG	Cellulose biosynthesis protein BcsG	3831107	Stop codon (C→T)
ΦWec188	R188	yhaH	Inner membrane protein YhaH	3405873-3405920	Frame shift*
ΦWec189	R189	nfrB	bacteriophage N4 adsorption protein B	576511	Stop codon (C→T)
		ompA	Outer membrane protein A	1046632-1046640	Frame shift (CACCAAAAG→.)
ΦWec190	R190	waaG	UDP-glucose:(heptosyl)LPS alpha-1,3-glucosyltransferase	3940586	Stop codon (A→T)
ΦWec191	R191	nfrB	bacteriophage N4 adsorption protein B	576613	Asp→Ala (T→G)
		tolA	TolA	767756	Stop codon (C→T)
ΦWec193	R193	waaY	LPS core heptose(II) kinase RfaY	3936259	Glu→Gly (C→A)
ΦWec196	R196	waaY	LPS core heptose(II) kinase RfaY	3936259	Glu→Gly (C→A)

*: CGTGGTGGGCAGTTCAGTTCCGCCGCCTGCACGACACCGACCGCTCGG→.
. means missing

136 **Efficiency of plating (EoP) of phages against target receptor mutant strains and Keio**
137 **collection**
138 The EoP of each phage against TK001 and target receptor mutant strains was evaluated
139 (Table 2). A decrease in EoP to less than 10-fold ($\log(\mathrm{EoP}) < -1$) was defined as a
140 reduction in infection efficiency. All phages could infect *bcsG* mutant (R187), a mutation
141 common to all strains; therefore, *bcsG* was not considered to be involved in target
142 recognition.
143 Regarding LPS synthesis genes, none of the phages could infect the *waaG* mutant (R190).
144 ΦWec172, 174, and 177 lost their ability to infect *waaW* mutant (R174) and *waaV*
145 mutants (R172, R177-1, and R177-2). ΦWec174, 177, 187, 189, 191, 193, and 196 could
146 not infect *waaY* mutants (R193 and R196). ΦWec179, 181, 186, 188, and 190 lost
147 infectivity against *waaT* and *manB-1* double mutant (R181).
148 For membrane protein and flagella genes, ΦWec179, 181, and 186 could not infect *fihD*
149 mutant (R186). ΦWec188 and 190 lost infectivity against *yhaH* mutant (R188). ΦWec189,
150 191, 193, and 196 could not infect *nfrB* and *ompA* double mutant (R189) and *nfrB* and
151 *tolA* double mutant (R191).
152 Using ΦWec187, 189, 191, 193, and 196, which can infect the K-12-derived BW25113
153 strain (wild-type of the Keio collection), we investigated their ability to infect Keio
154 collection strains with single gene deletions corresponding to the mutations found in
155 phage-resistant strains (Table S1). ΦWec187, 189, 191, 193, and 196 all failed to infect
156 Δ*nfrB*, Δ*waaG*, and Δ*waaF* strains. Additionally, ΦWec189, 191, 193, and 196 could not
157 infect the Δ*nfrA* strain, which is genetically redundant with *nfrB*.
158 Based on these results, the genes necessary for infection by each phage were classified
159 into six types (Table 3). Types a and b were very similar, differing only in the presence or

absence of a *waaY* mutation.

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

	waaG	*waaV*	*waaV*	*waaV*	*waaW*	*waaY*	*waaY*	*waaT*	*fihD*	*yhaH*	*nfrB*	*nfrB*	
				wecA				*manB-1*	*bcsG*		*ompA*	*tolA*	*bcsG*
Log(EoP)	R190	R172	R177-1	R177-2	R174	R193	R196	R181	R186	R188	R189	R191	R187
ΦWec187	ND	-0.2	-0.2	0.0	-0.4	ND	ND	-0.4	-0.2	0.0	0.0	-0.8	0.0
ΦWec189	ND	0.1	0.0	-0.1	0.0	ND	ND	-0.2	0.1	0.0	-2.2	ND	0.1
ΦWec191	ND	0.0	0.1	0.0	-0.1	ND	ND	0.1	0.0	0.3	-2.0	ND	0.4
ΦWec193	ND	0.1	0.0	0.0	0.0	ND	ND	0.1	0.2	0.3	-2.3	ND	0.2
ΦWec196	ND	0.1	0.2	0.1	0.1	ND	ND	0.3	0.3	0.3	-2.0	ND	0.1
ΦWec177	ND	ND	ND	ND	ND	ND	ND	-0.2	-0.9	-0.3	-0.7	-0.7	-0.3
ΦWec174	ND	ND	ND	ND	ND	ND	ND	-0.1	-0.5	0.1	-0.4	-0.3	-0.2
ΦWec172	ND	ND	ND	ND	ND	-0.3	-0.4	-0.6	-0.3	0.0	-0.4	-0.5	-0.1
ΦWec179	ND	0.2	-0.1	-0.1	0.0	-0.1	-0.1	-7.0	ND	0.1	0.0	0.0	-1.0
ΦWec181	ND	-0.2	-0.2	0.0	-0.2	0.1	0.1	-5.6	ND	-0.1	0.1	-0.1	-0.9
ΦWec186	ND	0.1	0.0	0.0	0.1	0.3	0.2	-4.7	ND	0.2	0.2	0.2	-0.8
ΦWec188	ND	0.0	0.0	0.0	-0.2	0.2	0.0	ND	0.1	ND	0.3	0.0	-0.1
ΦWec190	ND	0.2	0.1	0.2	0.1	0.3	0.1	ND	0.1	ND	0.1	0.1	0.2

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.

Table S1 Infectivity of K-12-infecting phages against Keio collection strains

Log(EoP)	*ΔnfrB*	*ΔompA*	*ΔnfrA*	*ΔtolA*	*ΔtolB*	*ΔwaaG*	*ΔwaaF*	*ΔwaaR*	*ΔwaaO*	*ΔmanB-1*	*ΔflhD*	*ΔyhaH*	*ΔbcsG*
ΦWec187	-1.9	0.0	-0.5	-0.3	-0.1	ND	ND	-0.1	-0.3	-0.1	-0.1	-0.1	0.1
ΦWec189	-3.4	-0.2	-3.3	0.1	0.0	ND	ND	-0.1	-0.1	0.0	0.1	-0.1	-0.2
ΦWec191	-1.2	-0.2	-1.3	-0.2	0.0	ND	ND	0.1	-0.1	0.1	0.0	-0.2	-0.1
ΦWec193	-3.6	-0.3	-3.3	-0.2	-0.3	ND	ND	-0.2	-0.2	-0.1	-0.2	-0.2	-0.3
ΦWec196	-3.6	0.1	-4.2	-0.3	0.1	ND	ND	0.0	-0.8	0.0	-0.2	-0.1	0.1

Table 3 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	

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

171 **Comparison of target receptors with various classification methods**

172 Our previous report suggested the possibility that grouping phages based on physiological
173 characteristics (using principal component analysis) could effectively cluster phages into
174 target receptors groupings/clusters. In this study, we conducted grouping based on
175 physiological characteristics, tail fiber phylogenetics, and phylogenetic analysis of whole
176 phage genomes, dividing the phages into three or five groups (Fig. 1, Table S2). When
177 grouped into three clusters, ΦWec179, 181, and 186 were classified differently depending
178 on the classification method (Fig. 1, Table S2). However, when divided into five clusters,
179 these phages were consistently grouped together regardless of the classification method
180 used.

181 The similarity between these groupings and the phage groups based on genes necessary
182 for infection was evaluated using four methods (Adjusted Rand Index (ARI), Normalized
183 Mutual Information (NMI), V-measure, and Fowlkes-Mallows Index). All evaluation
184 methods showed positive values for similarity with the classification based on target
185 receptors (Table 4).

186

187 The classification based on physiological characteristics and tail fibers (both divided into
188 three clusters) showed high similarity (ARI = 0.73) (Table 5). The classification based on
189 whole genomes (divided into three clusters) showed low to moderate similarity with other
190 classifications (ARI = 0.33-0.60), while classifications divided into five clusters generally
191 showed high similarity (ARI = 0.55-1.0), except for whole genome-based classification
192 which showed lower similarity (ARI = 0.35).

193

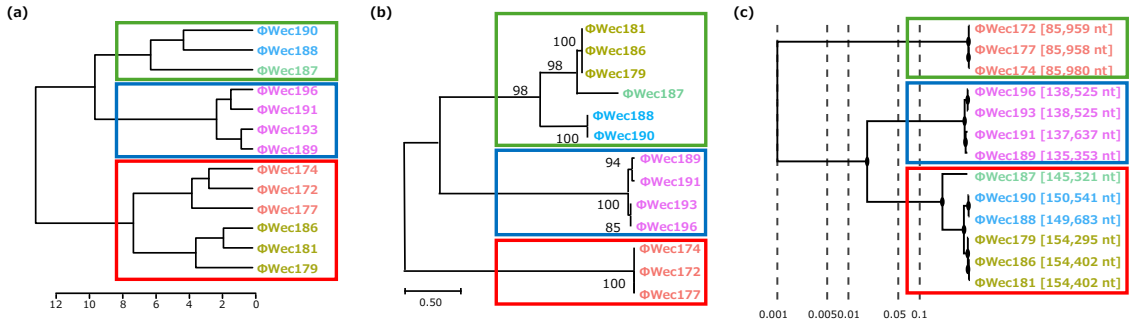


Fig. 1 Classification of phages using various methods

(a) Grouping based on physiological characteristics: lysis onset time, lysis duration time, lytic spectrum against 17 bacterial strains, adsorption rate constant, burst size, and phage yield after 1-day incubation. (b) Phylogenetic tree based on tail fiber genes. (c) Phylogenetic tree based on whole genomes. Phages enclosed in green/blue/red squares indicate that they are classified into the same cluster when divided into three groups. Additionally, phage names of the same color indicate that they belong to the same group when divided into five sub-groups in that grouping. The phylogenetic tree in (a) was drawn following the statistical analysis conducted in our previous paper [9] without any modifications.

Table S2 Classification results by each grouping

	Physiological cluster		Tail fiber cluster		Whole genome cluster	
	divided into three clusters	divided into five clusters	divided into three clusters	divided into five clusters	divided into three clusters	divided into five clusters
ΦWec172	P3-3	P5-4	T3-3	T5-5	W3-1	W5-1
ΦWec174	P3-3	P5-4	T3-3	T5-5	W3-1	W5-1
ΦWec177	P3-3	P5-4	T3-3	T5-5	W3-1	W5-1
ΦWec179	P3-3	P5-5	T3-1	T5-1	W3-3	W5-5
ΦWec181	P3-3	P5-5	T3-1	T5-1	W3-3	W5-5
ΦWec186	P3-3	P5-5	T3-1	T5-1	W3-3	W5-5
ΦWec187	P3-1	P5-2	T3-1	T5-2	W3-3	W5-3
ΦWec188	P3-1	P5-1	T3-1	T5-3	W3-3	W5-4
ΦWec189	P3-2	P5-3	T3-2	T5-4	W3-2	W5-2
ΦWec190	P3-1	P5-1	T3-1	T5-3	W3-3	W5-4
ΦWec191	P3-2	P5-3	T3-2	T5-4	W3-2	W5-2
ΦWec193	P3-2	P5-3	T3-2	T5-4	W3-2	W5-2
ΦWec196	P3-2	P5-3	T3-2	T5-4	W3-2	W5-2

Table 4 Similarity between various grouping methods and target receptor based grouping

Method	ARI	NMI	V-measure	Fowlkes-Mallows
P3	0.452327	0.709485	0.709485	0.72973
P5	0.730769	0.841091	0.841091	0.866667
T3	0.416051	0.64639	0.64639	0.703226
T5	0.730769	0.841091	0.841091	0.866667
W3	0.667742	0.821953	0.821953	0.833333
W5	0.730769	0.841091	0.841091	0.866667

Note: P, T and W indicate grouping based on physiological characteristics, tail fiber phylogeny, and whole genome phylogeny of phages, respectively, and the numbers 3 and 5 following these letters indicate division into 3 or 5 groups, respectively.

Table 5 Similarity between various methods

	P3	P5	T3	T5	W3	W5
P3		0.5882	0.7321	0.5882	0.6000	0.5882
P5			0.5455	1.0000	0.3529	1.0000
T3				0.5455	0.3304	0.5455
T5					0.3529	1.0000
W3						0.3529
W5						

Note: P, T and W indicate grouping based on physiological characteristics, tail fiber phylogeny, and whole genome phylogeny of phages, respectively, and the numbers 3 and 5 following these letters indicate division into 3 or 5 groups, respectively.

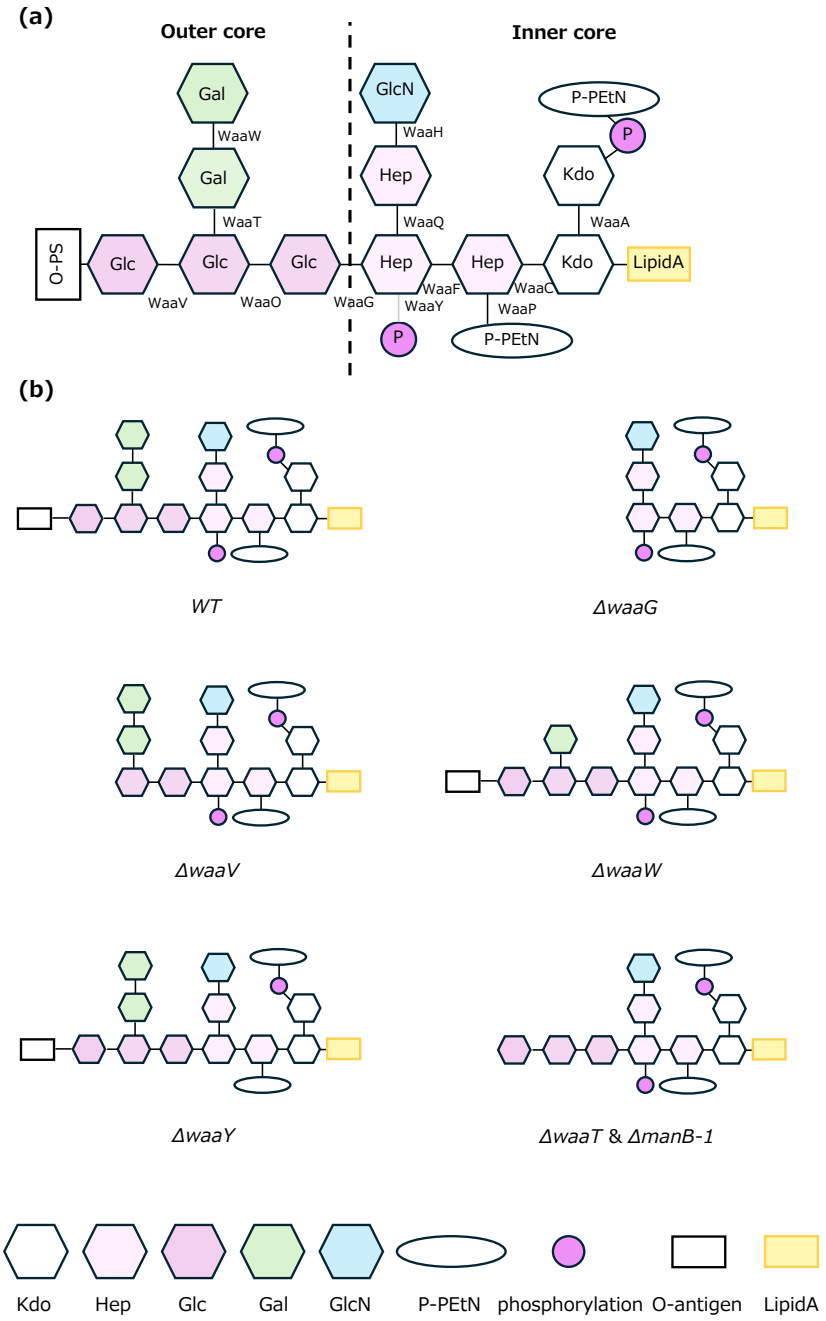
225 **LPS structure of TK001 wild type and phage resistant strains**

226 Mutations in genes involved in the LPS synthesis pathway were identified in many phage-
227 resistant *E. coli* strains. Most of the LPS-related enzymes mutated in the target receptor
228 mutant strains were associated with R-core synthesis, while *wecA* and *manB-1* were genes
229 for synthesizing sugars that constitute the O-antigen. Each *E. coli* strain possesses one of
230 five R-core types, and the analysis suggested that the R-core of TK001 was of the R1 type
231 (Fig. 2a)[10]. Variations in the R-core structure were suggested to arise from deficiencies in
232 the R-core synthesis pathway genes (Fig. 2b). The KEGG mapper results indicated the
233 presence of both *waaR* and *waaT* genes, as these genes share the same KEGG Orthology
234 (KO) identifier in the database. However, examination of the actual genome annotation
235 revealed that TK001 possesses only the *waaT* gene and not *waaR*, suggesting that the
236 cascade beyond *waaR* is not active.

237 DOC-PAGE analysis of LPS structure revealed that TK001 is similar to *E. coli* O111:B4,
238 suggesting the presence of LPS with a relatively compact O-antigen repeat unit in addition
239 to the R-core (Fig. 3a). Strains with mutations in genes unrelated to LPS synthesis showed
240 the same band pattern as TK001, indicating no change in LPS structure (Fig. 3b). In *waaV*
241 mutants (R172, R177-1, and R177-2), the upper band, O-antigen, completely disappeared.
242 The *waaW* mutant (R174) showed changes in the shape of the upper band and a downward
243 shift in the R-core band position. The *waaT* mutant (R181) showed complete
244 disappearance of the upper band and a downward shift in the R-core band position. The
245 *waaG* mutant (R190) showed near disappearance of the upper band and separation of the
246 R-core band into two. No clear changes in band pattern were observed in *waaY* mutants
247 (R193 and R196).

248

249



250

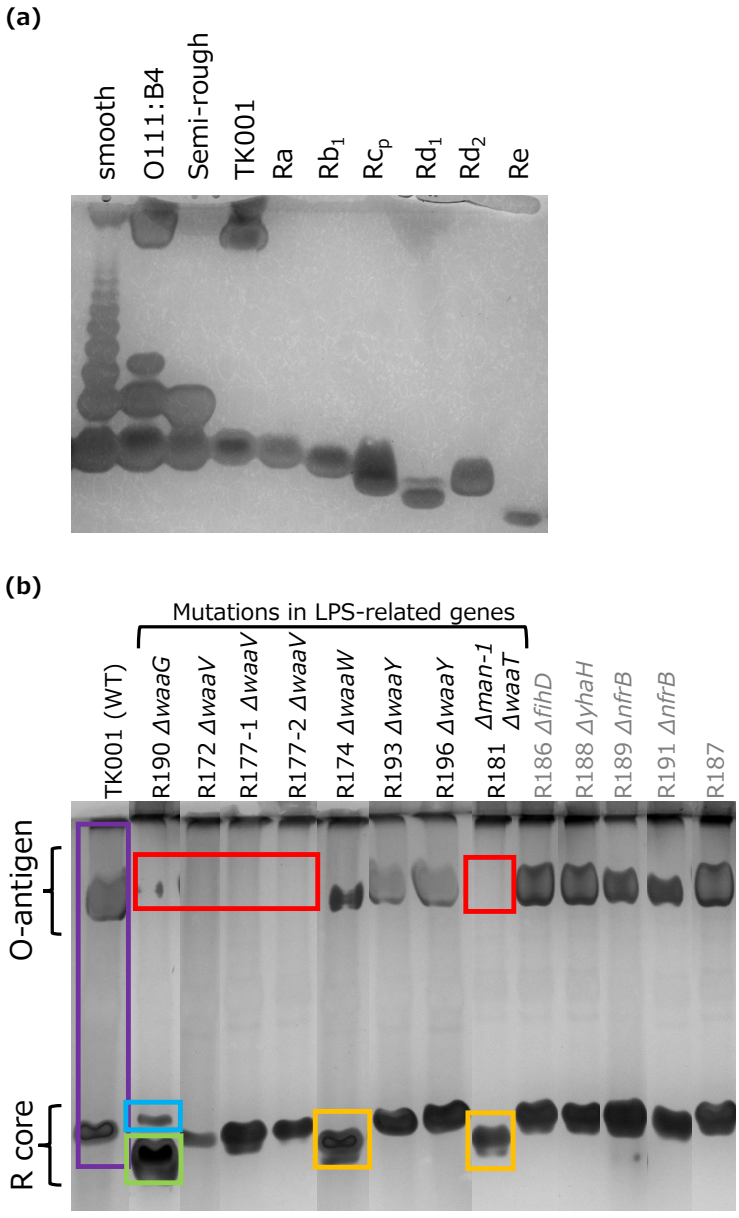
251 **Fig. 2 Prediction of LPS synthesis pathway in *Escherichia coli* TK001 strain by**

252 **KEGG mapping**

253 (a) R-core structure and related synthesis enzymes. (b) Predicted LPS structures of LPS-

254 related gene mutants.

255



256

257 **Fig. 3 Phenotypic confirmation of LPS extracted from each strain by DOC-PAGE**

258 (a) Comparison of TK001 strain with known strains. "Smooth" is LPS from *Salmonella*

259 *enterica* serovar Typhimurium, showing ladder-like bands characteristic of smooth strains

260 with varying numbers of repeats of O-antigen. "O111:B4" is LPS from *Escherichia coli*

261 O111:B4, showing a band pattern characteristic of smooth strains with relatively compact

262 long repeat units. "Semi-rough" is LPS from *S. enterica* serovar Typhimurium SL1034

263 (semi-rough chemotype). “Ra” is Ra chemotype. “Rb” is Rb chemotype. “Rc” is Rc
264 chemotype. Purple highlights indicate the characteristic band pattern of TK001,
265 consisting of upper and lower bands. Red highlights indicate regions where the upper
266 band is absent. Orange and green highlights show the downward shift of the lower bands
267 compared to the wild type. Blue highlights indicate the presence of an additional band
268 distinct from the band marked by the green highlight. (b) Comparison between mutant
269 strains. The order has been rearranged for ease of viewing. Purple highlights indicate the
270 characteristic band pattern of TK001, consisting of upper and lower bands. Red highlights
271 indicate regions where the upper band is absent. Orange and green highlights show the
272 downward shift of the lower bands compared to the wild type. Blue highlights indicate
273 the presence of an additional band distinct from the band marked by the green highlight.
274

275 **Discussion**

276 While phage therapy using a cocktail of phages targeting different receptors is known to
277 be effective, methods to identify phage target receptors remain limited. Moreover, current
278 receptor identification methods are time consuming and cannot be scaled. In our previous
279 study, we demonstrated that effective phage cocktails could be prepared by grouping
280 phages based on their physiological characteristics, but the actual receptors of the phages
281 were not confirmed. In this study, we validated our previous clustering method based on
282 physiological characteristics of *E. coli* phages through detailed analysis of each phage's
283 target receptor.

284 Whole genome analysis revealed the LPS structure of the TK001 strain, and mutation
285 analysis of phage-resistant strains identified mutations in genes encoding flagellar
286 proteins, outer or inner membrane proteins, or LPS synthesis enzymes involved in the

structures recognized by each phage.

Furthermore, comparing classification methods based on physiological characteristics, tail fiber phylogeny, and whole genome phylogeny showed that all methods exhibited high correlation with target receptors when grouped into five clusters (ARI: 0.731). These results suggest that phage classification based on phenotype or genotype reflects target receptor specificity, supporting the validity of clustering based on physiological characteristics as an effective method for reflecting differences in target receptors.

Phage classification methods

In this study, we compared three approaches for phage classification: physiological characteristics, tail fiber phylogeny, and whole genome phylogeny. Comparison of these methods with grouping based on target receptors showed that all methods performed better than random classification (ARI > 0). Notably, when grouped into five clusters, all methods showed high agreement (ARI: 0.731), indicating strong interchangeability among these classification methods.

This high level of agreement suggests a close relationship between phage phenotype (physiological characteristics), structural genes (tail fibers), and genetic background (whole genome). It is particularly noteworthy that clustering based on physiological characteristics showed high similarity with molecular phylogenetic classifications. This implies that phage phenotypes strongly reflect genotypes, suggesting the possibility of estimating genetic relationships of phages to some extent through relatively simple analysis of physiological characteristics.

While there was a high level of agreement when grouping the phages into five clusters, the agreement between methods decreased when phages were grouped into three clusters. In particular, differences were observed in the classification of ΦWec179, 181, and 186

between physiological characteristic-based grouping and other methods. This result suggests that each method may capture different aspects of phage characteristics. For instance, physiological characteristics reflect functional aspects of phages such as host range and growth dynamics, while tail fiber and whole genome analyses more directly reflect genetic relationships. Therefore, using these methods complementarily may be effective for a more comprehensive understanding of phage diversity.

Importantly, none of the methods completely matched the classification based on target receptors. This suggests that phage target receptor specificity is a complex characteristic that cannot be fully explained by a single classification criterion. The recognition of target receptors is likely determined by not only specific regions of phage tail fibers but also other structural proteins. Furthermore, subtle differences may exist that cannot be fully captured by whole genome analysis or physiological characteristics alone. For example, studies on bacteriophage T4 have revealed that structures such as the collar and whiskers act as environment-sensing devices, regulating the retraction of long tail fibers under unfavorable conditions and thus preventing infection, demonstrating that phage-host recognition can involve structural elements beyond tail fibers to modulate infection processes[11]. These results highlight the challenges in phage classification and functional prediction, and reflects the complexity of phage-host coevolution.

Nevertheless, this study demonstrated that clustering based on physiological characteristics is an effective method that reflects phage target receptor specificity to some extent. These findings contribute to the development of phage classification research and provide an important foundation for the development of effective phage therapy.

Phage target receptors

335 In this study, we were able to identify the target receptors of each phage in detail by
336 combining the analysis of phage-resistant strains with LPS structure analysis. These
337 results reveal the diversity and specificity of phage host recognition mechanisms and
338 provide new insights into the relationship between LPS structure and phage infectivity.
339 The results of LPS structure analysis clearly show the specific effects of each gene
340 mutation on LPS structure. The complete disappearance of the upper band in DOC-PAGE
341 for *waaV* mutants (R172, R177-1, and R177-2) strongly suggests that in the TK001 strain,
342 the O-antigen is attached to the Glcβ added by WaaV. Additionally, the downward shift
343 of the R-core bands in *waaW* and *waaT* mutants (R174 and R181) indicates that the
344 deficiency of these genes results in a shorter R-core length. Interestingly, in the *waaG*
345 mutant (R190), the upper band almost disappeared, and the R-core band separated into
346 two. Since *waaG* is an early gene in outer core synthesis, this result suggests that while
347 R-core synthesis was incomplete, some compensatory mechanism may have led to the
348 expression of a unique LPS. It is worth noting that mutations that shorten the R-core are
349 known to destabilize the outer membrane, increasing susceptibility to external factors
350 such as detergents and antimicrobial peptides. Gram-negative bacteria often activate
351 compensatory pathways, such as the PhoPQ and PmrAB systems, to modify lipid A or the
352 core oligosaccharide with charged non-stoichiometric substituents, thereby restoring
353 membrane stability and reducing permeability[12,13]. Furthermore, it is known that mutation
354 of a single LPS biosynthesis gene can lead to unexpected structural changes: Klein et al.
355 observed that *waaG* deletion in *E. coli* K-12 resulted in increased incorporation of a third
356 Kdo residue and rhamnose, leading to accumulation of different structures of
357 glycoforms[14]. This altered glycoform distribution affected LPS heterogeneity. These
358 findings align with reports that even minor changes in the R-core structure can result in

359 complex physiological adaptations to maintain outer membrane integrity and resistance
360 to external factors. On the other hand, no clear changes in band pattern were observed in
361 *waaY* mutants (R193 and R196), indicating that WaaY, which is involved only in Hep
362 phosphorylation, has little effect on the overall length of LPS. Notably, although
363 ΦWec172, 174, and 177 were suggested to be similar in our previous report [9] due to their
364 common isolation source and timing, similar lysis range, and over 99% homology in ANI-
365 based similarity evaluation (Table S3), ΦWec172 did not require WaaY's function, unlike
366 ΦWec174 and 177. These phages have two types of tail fibers: Tail fiber 1 consisting of
367 844 amino acid residues and Tail fiber 2 consisting of 390 amino acid residues.
368 Particularly noteworthy are the N777S mutation in Tail fiber 1 of ΦWec174 and the
369 K286E mutation in Tail fiber 2 of ΦWec177. These differences suggest that ΦWec172
370 may interact with LPS differently from the other two phages and may not require Heptose
371 phosphorylation by WaaY. This observation demonstrates the complexity and flexibility
372 of phage receptor recognition and may provide an example of adaptation strategies in
373 phage-host interactions. However, as no unique mutations were found in these tail fibers
374 for ΦWec172 alone, further genomic analysis, structural analysis, or mutation
375 experiments are necessary to elucidate the mechanism.
376 The changes in LPS structure were closely related to the infectivity of each phage. The
377 most important finding was that all phages were unable to infect the *waaG* mutant (R190,
378 clearly demonstrating that LPS is essential for infection by all phages used in this study.
379 Furthermore, interesting results were obtained regarding the target receptor recognition
380 of ΦWec172, 174, and 177. These phages could not infect the *waaW* mutant but could
381 infect the strain with a mutation in *waaT*, which functions just before WaaW in the LPS
382 synthesis pathway. In the LPS synthesis pathway, WaaT transfers Galactose, followed by

383 WaaW transferring second Galactose. This result suggests that these phages primarily
384 target the Galactose transferred by WaaW. However, the fact that they can also infect the
385 WaaT mutant suggests that similar structures, such as Glucose transferred by WaaO, may
386 serve as alternative targets. This flexible recognition mechanism might be one of the
387 phages' strategies for adapting to their hosts.
388 In conclusion, this study provides important insights into *E. coli* phage target receptors
389 (Fig. 4) and the relationship between various classification methods. Our findings reveal
390 that phage host recognition mechanisms are diverse and specific, with many phages
391 recognizing specific structures of LPS, while others target membrane proteins or flagella.
392 Notably, we demonstrated that subtle chemical modifications of LPS, such as Heptose
393 phosphorylation, are crucial for phage recognition. Furthermore, we found a high degree
394 of agreement between classifications based on previously reported physiological
395 characteristics, tail fiber phylogeny, and whole genome phylogeny when grouped into
396 five clusters, suggesting a close relationship between phage phenotypes, specific
397 structural genes, and overall genetic background. Through robust molecular analysis of
398 13 phages, including detailed characterization of their physiological features, whole
399 genome sequences, and target receptors, we have definitively validated our previously
400 proposed classification method by demonstrating its correlation with actual phage target
401 receptors. Our use of *E. coli* TK001, a strain that shows enhanced growth during colitis,
402 as a host strain further demonstrates the applicability of our approach to clinically relevant
403 contexts. While our findings are based on *E. coli* phages, further studies about phages
404 infecting other bacterial species are needed, in clinical settings where conventional
405 receptor identification is impractical and rapid decision-making is crucial, these validated
406 classification methods offer powerful tools for creating effective phage cocktails that can

407 suppress the emergence of phage-resistant bacteria, marking a significant advancement

408 in the field of phage therapy.

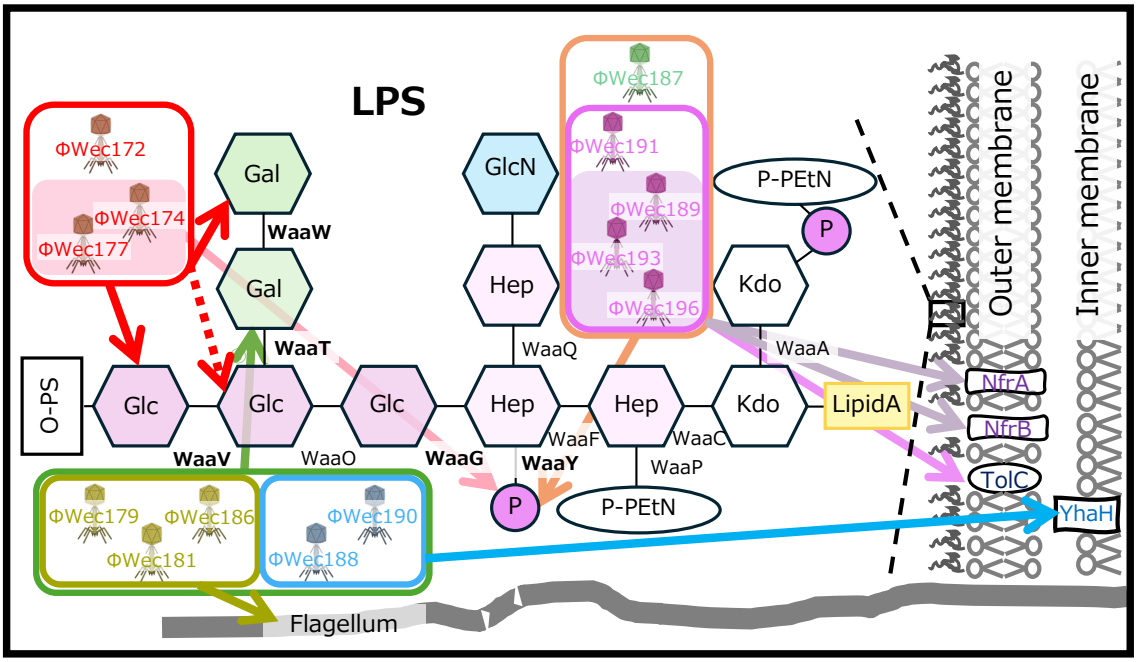
409

	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												

410

411 **Table S3 ANI of phages**

412



413

414 **Fig. 4 Target receptors of each phage**

415 Schematic representation of the target receptors for the studied phages. The diagram

416 illustrates the diversity of receptors including LPS components, membrane proteins, and

417 flagella. The KEGG mapper results indicated the presence of both *waaR* and *waaT* genes,
418 as these genes share the same KEGG Orthology (KO) identifier in the database. However,
419 examination of the actual genome annotation revealed that TK001 possesses only the
420 *waaT* gene and not *waaR*. Boxed phages represent groups classified by physiological
421 characteristics, whole genome, or tail fiber phylogeny. Arrows extending from each phage
422 indicate their respective target receptors.
423

424 **Materials and methods**

425 **Reagents and bacteria**

426 Luria–Bertani (LB) medium (BD Difco, Franklin Lakes, NJ, USA) and the soft agar
427 containing 0.5% agar were used for bacterial and phage cultivation. *E. coli* strain TK001,
428 originally isolated from dextran sodium sulfate (MP Biomedicals, Soho, OH, USA)-
429 induced colitis mouse feces, was used as the host for phage screening in this study[9].

430 **Phage preparation**

431 Thirteen phages were propagated using *E. coli* TK001 as the host through the plate lysate
432 method. Briefly, 100 μL of overnight-cultured TK001 and 100 μL of phage suspension
433 (10^4 plaque-forming unit (PFU)/mL) were added to soft agar (0.5% agar) and overlaid on
434 agar plates. After overnight incubation at 37 °C, 5 mL of SM buffer (0.1 M NaCl, 8 mM
435 $MgSO_4 \cdot 7H_2O$, 50 mM Tris-HCl buffer [pH=7.5], 0.1% gelatin) was added to each plate
436 and subjected to orbital shaking (approximately 250 rpm, 2-3 hours). The supernatant was
437 centrifuged (9,000 rpm, 15 min, 4 °C) to separate solid and liquid components. The
438 supernatant was transferred to a new tube, and 1 mL of chloroform was added. After
439 vortexing and settling, and purified phage was decanted and served as the phage stock.

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

441 Phage-resistant strains were generated using both liquid and solid culture methods:
442 *Liquid culture method*: Log-phase TK001 culture was diluted 100-fold, and 100 μL of
443 high-titer phage suspension (>10^8 PFU/mL, MOI > 10) was added. The mixture was
444 incubated at 37°C with shaking at 40 rpm in a compact rocking incubator (TVS062CA,
445 Advantec Toyo Co. Ltd., Tokyo, Japan). The emergence of resistant bacteria was
446 confirmed by turbidity curves the following day. The culture was centrifuged (9000 rpm,
447 5 min, 4 °C), and the pellet was washed with 1× phosphate-buffered saline (PBS). The
448 pellet was serially diluted in 1× PBS (1 mL), and 20 μL of each dilution was spotted onto
449 soft agar containing 100 μL of high-titer phage. Eight phage-resistant *E. coli* colonies
450 were isolated for each phage and purified by streak culture at least twice. Purified colonies
451 were inoculated into 1 mL of LB, cultured for 11 hours, and stored at -80 °C.
452 *Solid culture method*: 100 μL of overnight-cultured TK001 and 100 μL of high-titer phage
453 suspension (>10^8 PFU/mL) were added to soft agar and overlaid on agar plates. Four
454 colonies per phage (one for ΦWec190 and three for ΦWec193) were isolated the
455 following day and purified by streak culture at least twice. Purified colonies were
456 processed and stored as described in the liquid culture method.
457 Phage adsorption assay was used to select target receptor mutants from the acquired
458 phage-resistant strains. A mixture of overnight-cultured resistant strain (100 μL) and LB
459 (890 μL) was combined with phage suspension (>10^7 PFU/mL, 100 μL). Samples (10 μL)
460 were taken at 0- and 10-min post-phage addition, mixed with 990 μL SM buffer (pre-
461 added with 100 μL chloroform), and centrifuged (15,000 rpm, >2 min, 4 °C). The
462 supernatant (100 μL) and overnight-cultured TK001 (100 μL) were mixed in soft agar
463 and overlaid on agar plates. After overnight incubation, the unadsorbed ratio was
464 calculated by dividing the number of plaques at 10 min by the number at 0 min. Strains

465 with an adsorption rate (100 - unadsorbed ratio) below 20% were designated as target
466 receptor mutants.

467 **Spot test using target receptor mutants**

468 Each phage was serially diluted 10-fold up to 10^{-7} and spotted (1 μL) onto target
469 receptor mutant lawns[15]. The efficiency of plating (EoP) of the mutants relative to the
470 wild type was calculated. Biological replicates were performed 3-7 times.

471 **Spot test using the Keio collection**

472 For the five *E. coli* phages (ΦWec187, 189, 191, 193, and 196) capable of infecting K-12
473 strains[9], spot tests were conducted using the Keio collection (NBRP, Japan), a
474 comprehensive gene knockout library[16,17]. The EoP of Keio collection strains relative to
475 the wild-type BW25113 was calculated. The Keio collection strains used were limited to
476 those predicted to be related to the target receptors of ΦWec187, 189, 191, 193, and 196
477 based on phage-resistant TK001 (Table S1). Biological replicates were performed 3 times.

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

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

482 Sequencing was performed by BGI Genomics. Continuous genome sequencing was
483 assembled using two different sequencing technologies. Shotgun sequencing was
484 performed using DNBSEQ G400 (MGI) with a read length of 150 bp. Clean data was
485 generated using SOAPnuke (BGI) to remove adapters and low-quality reads (parameters:
486 "-n 0.01 -l 20 -q 0.4 --adaMis 3 --outQualSys 1")[18]. Long reads obtained by Oxford
487 Nanopore Technologies were processed using porechop (ver. 0.2.4, default parameters).
488 The processed shotgun and long-read data were assembled using SPAdes (ver. 3.15.5,

489 parameters: "-k auto --cov-cutoff auto --careful -t 12") for hybrid assembly[19]. Annotation
490 of the resulting contigs was performed using prokka[20]. Basic statistical analysis of the
491 contigs was conducted using seqkit[21].

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

493 DNA was extracted from selected target receptor mutant strains as described for the wild-
494 type strain. BGI performed shotgun sequencing and read processing similarly to the wild-
495 type strain. The mutant strain shotgun reads were mapped to the indexed wild-type
496 sequence using bwa. Samtools was used to convert the mapping results to BAM format
497 and sort them. The sorted BAM files were indexed using samtools. Guided assembly was
498 performed using pilon (ver. 1.24, parameters: "--vcf --tracks --changes --verbose") with
499 the index files[22]. Non-synonymous mutations were detected by comparing each resistant
500 strain's sequence obtained through guided assembly with the wild-type strain using
501 nucdiff[23].

502 **Preparation of dried defatted bacterial cells**

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

508 The cells were then defatted through sequential washing steps. First, cells were
509 resuspended in 30 mL of 100% ethanol using a spatula, followed by centrifugation (4,800
510 rpm, 15 min, 4°C). This ethanol washing step was repeated until the cell pellet lost its
511 sticky consistency (typically three times), with the supernatant turning yellow. The pellet
512 was then washed once with 30 mL acetone, using gentle deceleration during

513 centrifugation to prevent pellet disruption. Finally, cells were washed with 30 mL of
514 hexane or diethyl ether.
515 The resulting pellet was dried under vacuum in a desiccator for 1-3 days. This process
516 typically yielded 4-5 g of off-white to white dried defatted bacterial cells from *E. coli*
517 cultures.

518 **LPS structure analysis**

519 LPS structure analysis of wild type and target receptor mutant strains was performed
520 using two methods: pathway analysis to predict the LPS synthesis pathway from retained
521 genes, and sodium deoxycholate-polyacrylamide gel electrophoresis (DOC-PAGE) for
522 LPS structure analysis.
523 *Pathway analysis*: KEGG Mapper was used to map retained genes to known synthesis
524 pathways based on Kegg Orthology (KO) assignment results obtained using
525 BlastKOALA[24,25,26].
526 *Structural analysis*: To obtain LPS from bacterial cells, dried defatted bacterial cells were
527 first prepared. LPS extraction from dried defatted bacterial cells was performed using the
528 hot phenol-water extraction method[27]. Extracted LPS was visualized by DOC-PAGE
529 followed by silver staining[28]. Silver staining was performed using the Silver Stain MS Kit
530 (Fujifilm Wako, 299-58901, Osaka, Japan), with a modified sodium periodate oxidation
531 step between first and second fixative solutions[29]. The gel was shaken in 10 times its
532 volume of 1% w/v sodium periodate solution for 20 min, followed by three 5-min washes
533 with MilliQ water. As control samples for the LPS structure of TK001, LPS extracted
534 from the following strains was used: *Salmonella enterica* serovar Typhimurium (Sigma;
535 catalogue no. L6511), *Escherichia coli* O111:B4 (Sigma; catalogue no. L4130), and
536 various *S. enterica* serovars with different chemotypes (LPSs were extracted for bacteria

537 of kindly provided by Dr. Hajime Ikigai).

538 **Principal component analysis (PCA) and clustering**

539 PCA and clustering were performed as previously reported[9].

540 **Whole genome construction of phages**

541 Sequencing data were assembled using Velvet software (ver. 1.2.10)[30].

542 **Phylogenetic tree based on tail fibers**

543 Whole genome information of each phage was annotated using RAST, and tail fiber
544 sequences were extracted[31]. Additionally, tail fibers of the closest relatives were obtained
545 by BLAST search of each phage's whole genome, and similar sequences were extracted
546 from the whole genomes of phages drawing these sequences. The obtained sequences
547 were translated into amino acids, and for each phage, the sequences were artificially
548 combined into a single sequence. A phylogenetic tree was constructed using MEGA X
549 with the Maximum-likelihood method and 1,000 bootstrap replicates[32].

550 **Whole genome phylogenetic tree and similarity**

551 The whole genome phylogenetic tree was created using Viptree[33]. The Average
552 Nucleotide Identity (ANI) was calculated using pyani[34]. The parameters used were: pyani:
553 anib and ANI calculator: minimum identity 0%, with other parameters set to default.

554 **Comparison of similarity between methods**

555 Data analysis was performed using Python 3.9.12, with major packages including pandas
556 (1.4.2), numpy (1.22.3), and scikit-learn (1.0.2). The similarity of clustering results was
557 evaluated using Adjusted Rand Index (ARI), Normalized Mutual Information (NMI), V-
558 measure, and Fowlkes-Mallows Index. Label normalization was performed in the
559 similarity evaluation to improve the accuracy of comparison between methods. These
560 analyses were executed on Linux.

561 **Statistical analysis**

562 The data were analyzed using R (ver. 4.0.3) and its packages[35,36,37,38,39,40,41,42,43,44,45,46].

563

564 **Acknowledgements**

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

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

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

575 We would like to thank Matthew Imanaka for English language editing.

576

577 **FUNDING INFORMATION**

578 This research was supported by a Waseda University Grant for Special Research Projects
579 (Project number: 2022C-160).

580

581 **References**

582 1. Chen, P. *et al.* LamB, OmpC, and the Core Lipopolysaccharide of *Escherichia coli*
583 K-12 Function as Receptors of Bacteriophage Bp7. *J. Virol.* **94**, 1–14 (2020).

584 2. Tanji, Y. *et al.* Toward rational control of *Escherichia coli* O157:H7 by a phage

585 cocktail. *Appl. Microbiol. Biotechnol.* **64**, 270–274 (2004).

586 3. Yang, Y. *et al.* Development of a Bacteriophage Cocktail to Constrain the
587 Emergence of Phage-Resistant *Pseudomonas aeruginosa*. *Front. Microbiol.* **11**, 1–
588 12 (2020).

589 4. Tanji, Y. *et al.* Therapeutic use of phage cocktail for controlling *Escherichia coli*
590 O157:H7 in gastrointestinal tract of mice. *J. Biosci. Bioeng.* **100**, 280–287 (2005).

591 5. Bai, J., Jeon, B. & Ryu, S. Effective inhibition of *Salmonella Typhimurium* in fresh
592 produce by a phage cocktail targeting multiple host receptors. *Food Microbiol.* **77**,
593 52–60 (2019).

594 6. Schooley, R. T. *et al.* Development and use of personalized bacteriophage-based
595 therapeutic cocktails to treat a patient with a disseminated resistant *Acinetobacter*
596 *baumannii* infection. *Antimicrob. Agents Chemother.* **61**, e00954-17 (2017).

597 7. Nakamura, K. *et al.* Fluctuating Bacteriophage-induced *galU* Deficiency Region
598 is Involved in Trade-off Effects on the Phage and Fluoroquinolone Sensitivity in
599 *Pseudomonas aeruginosa*. *Virus Res.* **306**, 198596 (2021).

600 8. Liang, J. *et al.* DTDP-rhamnosyl transferase RfbF, is a newfound receptor-related
601 regulatory protein for phage phiYe-F10 specific for *Yersinia enterocolitica*
602 serotype O:3. *Sci. Rep.* **6**, 1–11 (2016).

603 9. Kaneko, T., Osaka, T. & Tsuneda, S. Tailoring Effective Phage Cocktails for
604 Long-Term Lysis of *Escherichia coli* Based on Physiological Properties of
605 Constituent Phages. *PHAGE* **4**, 128–135 (2023).

606 10. Vinogradov, E. V. *et al.* The structures of the carbohydrate backbones of the
607 lipopolysaccharides from *Escherichia coli* rough mutants F470 (R1 core type) and
608 F576 (R2 core type). *Eur. J. Biochem.* **261**, 629–639 (1999).

609 11. Fokine, A. *et al.* The Molecular Architecture of the Bacteriophage T4 Neck. *J. Mol.*
610 *Biol.* **23**, 1–7 (2014).
611 12. Simpson, B. W. & Trent, M. S. Pushing the envelope: LPS modifications and their
612 consequences. *Nat. Rev. Microbiol.* **17**, 403–416 (2019).
613 13. Qian, J., Garrett, T. A. & Raetz, C. R. H. In vitro assembly of the outer core of the
614 lipopolysaccharide from *Escherichia coli* K-12 and *Salmonella typhimurium*.
615 *Biochemistry* **53**, 1250–1262 (2014).
616 14. Klein, G., Lindner, B., Brade, H. & Raina, S. Molecular basis of
617 lipopolysaccharide heterogeneity in *Escherichia coli*: Envelope stress-responsive
618 regulators control the incorporation of glycoforms with a third 3-deoxy-α-D-
619 manno-oct-2-ulosonic acid and rhamnose. *J. Biol. Chem.* **286**, 42787–42807
620 (2011).
621 15. Mattila, S., Ruotsalainen, P. & Jalasvuori, M. On-Demand Isolation of
622 Bacteriophages Against Drug-Resistant Bacteria for Personalized Phage Therapy
623 Bacteria Strains and Culturing. *Front. Microbiol.* **6**, 1–7 (2015).
624 16. Baba, T. *et al.* Construction of *Escherichia coli* K-12 in-frame, single-gene
625 knockout mutants : the Keio collection. *Mol. Syst. Biol.* 2006.0008 (2006)
626 doi:10.1038/msb4100050.
627 17. Yamamoto, N. *et al.* Update on the Keio collection of *Escherichia coli* single-gene
628 deletion mutants. *Mol. Syst. Biol.* **5**, (2009).
629 18. Chen, Y. *et al.* SOAPnuke: A MapReduce acceleration-supported software for
630 integrated quality control and preprocessing of high-throughput sequencing data.
631 *Gigascience* **7**, 1–6 (2018).
632 19. Bankevich, A. *et al.* SPAdes: A new genome assembly algorithm and its

633 applications to single-cell sequencing. *J. Comput. Biol.* **19**, 455–477 (2012).
634 20. Seemann, T. Prokka: Rapid prokaryotic genome annotation. *Bioinformatics* **30**,
635 2068–2069 (2014).
636 21. Shen, W., Le, S., Li, Y. & Hu, F. SeqKit: A cross-platform and ultrafast toolkit for
637 FASTA/Q file manipulation. *PLoS One* **11**, 1–10 (2016).
638 22. Walker, B. J. *et al.* Pilon: An integrated tool for comprehensive microbial variant
639 detection and genome assembly improvement. *PLoS One* **9**, (2014).
640 23. Khelik, K., Lagesen, K., Sandve, G. K., Rognes, T. & Nederbragt, A. J. NucDiff:
641 In-depth characterization and annotation of differences between two sets of DNA
642 sequences. *BMC Bioinformatics* **18**, 1–14 (2017).
643 24. Kanehisa, M., Sato, Y. & Morishima, K. BlastKOALA and GhostKOALA: KEGG
644 Tools for Functional Characterization of Genome and Metagenome Sequences. *J.*
645 *Mol. Biol.* **428**, 726–731 (2016).
646 25. Kanehisa, M. & Sato, Y. KEGG Mapper for inferring cellular functions from
647 protein sequences. *Protein Sci.* **29**, 28–35 (2020).
648 26. Kanehisa, M., Sato, Y. & Kawashima, M. KEGG mapping tools for uncovering
649 hidden features in biological data. *Protein Sci.* **31**, 47–53 (2022).
650 27. Westphal, O., Jann, K. & Himmelspach, K. Chemistry and immunochemistry of
651 bacterial lipopolysaccharides as cell wall antigens and endotoxins. *Prog. Allergy*
652 **33**, 9–39 (1983).
653 28. Kokubo, T. & Galanos, C. Analysis of *Salmonella* lipopolysaccharides by sodium
654 deoxycholate-polyacrylamide gel electrophoresis. *J. Chromatogr.* **450**, 381–387
655 (1988).
656 29. Fomsgaard, A., Freudenberg, M. A. & Galanos, C. Modification of the silver

657 staining technique to detect lipopolysaccharide in polyacrylamide gels. *J. Clin.*
658 *Microbiol.* **28**, 2627–2631 (1990).
659 30. Zerbino, D. R. & Birney, E. Velvet: Algorithms for de novo short read assembly
660 using de Bruijn graphs. *Genome Res.* **18**, 821–829 (2008).
661 31. Aziz, R. K. *et al.* The RAST Server: Rapid annotations using subsystems
662 technology. *BMC Genomics* **9**, 1–15 (2008).
663 32. Kumar, S., Stecher, G., Li, M., Knyaz, C. & Tamura, K. MEGA X: Molecular
664 evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* **35**,
665 1547–1549 (2018).
666 33. Nishimura, Y. *et al.* ViPTree: The viral proteomic tree server. *Bioinformatics* **33**,
667 2379–2380 (2017).
668 34. Pritchard, L., Glover, R. H., Humphris, S., Elphinstone, J. G. & Toth, I. K.
669 Genomics and taxonomy in diagnostics for food security: Soft-rotting
670 enterobacterial plant pathogens. *Anal. Methods* **8**, 12–24 (2016).
671 35. R Core Team. *R: A Language and Environment for Statistical Computing*. vol. 2
672 (R Foundation for Statistical Computing, 2017).
673 36. Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M. & Hornik, K. cluster: Cluster
674 Analysis Basics and Extensions. https://cran.r-project.org/package=cluster (2019).
675 37. Vincent Q, V. ggbiplot: A ggplot2 based biplot. https://github.com/vqv/ggbiplot
676 (2011).
677 38. Wickham, H. ggplot2: Elegant Graphics for Data Analysis. *Springer-Verlag New*
678 *York* https://ggplot2.tidyverse.org/ (2016).
679 39. Slowikowski, K. ggrepel: Automatically Position Non-Overlapping Text Labels
680 with 'ggplot2'. https://cran.r-project.org/package=ggrepel (2021).

40. Alboukadel, K. & Fabian, M. Extract and Visualize the Results of Multivariate Data Analyses [R package factoextra version 1.0.7]. https://cran.r-project.org/package=factoextra (2020).

41. Schauberger, P. & Walker, A. openxlsx: Read, Write and Edit xlsx Files. https://cran.r-project.org/package=openxlsx (2021).

42. Wickham, H. The split-apply-combine strategy for data analysis. *J. Stat. Softw.* **40**, 1–29 (2011).

43. Wickham, H. Reshaping Data with the reshape Package. *J. Stat. Softw.* **21**, 1–20 (2007).

44. Bache, M. S. & Wickham, H. magrittr: A Forward-Pipe Operator for R. R package version 2.0.1. https://cran.r-project.org/package=magrittr (2022).

45. Wickham, H. *et al.* Welcome to the Tidyverse. *J. Open Source Softw.* **4**, 1686 (2019).

46. Barret Schloerke *et al.* GGally: Extension to 'ggplot2'. https://cran.r-project.org/package=GGally (2021).