論文 / 著書情報
Article / Book Information

題目(和文)	
Title(English)	Genetic analysis of host-phage interaction of Escherichia coli O157:H7 and its specific phage, PP01
著者(和文)	星賀史也
Author(English)	Fumiya Hoshiga
出典(和文)	学位:博士(工学), 学位授与機関:東京工業大学, 報告番号:甲第11102号, 授与年月日:2019年3月26日, 学位の種別:課程博士, 審査員:丹治 保典,和地 正明,廣田 順二,小倉 俊一郎,八波 利恵
Citation(English)	Degree:Doctor (Engineering), Conferring organization: Tokyo Institute of Technology, Report number:甲第11102号, Conferred date:2019/3/26, Degree Type:Course doctor, Examiner:,,,,
学位種別(和文)	博士論文
Type(English)	Doctoral Thesis

Genetic analysis of host-phage interaction of *Escherichia coli* O157:H7 and its specific phage, PP01

Doctoral Thesis by

Fumiya Hoshiga

Department of Bioengineering

School of Life Science and Technology

Tokyo Institute of Technology

2019

Supervisor: Professor Yasunori Tanji

List of contents

List of tables

List of figures

CHAPTER 1 General Introduction

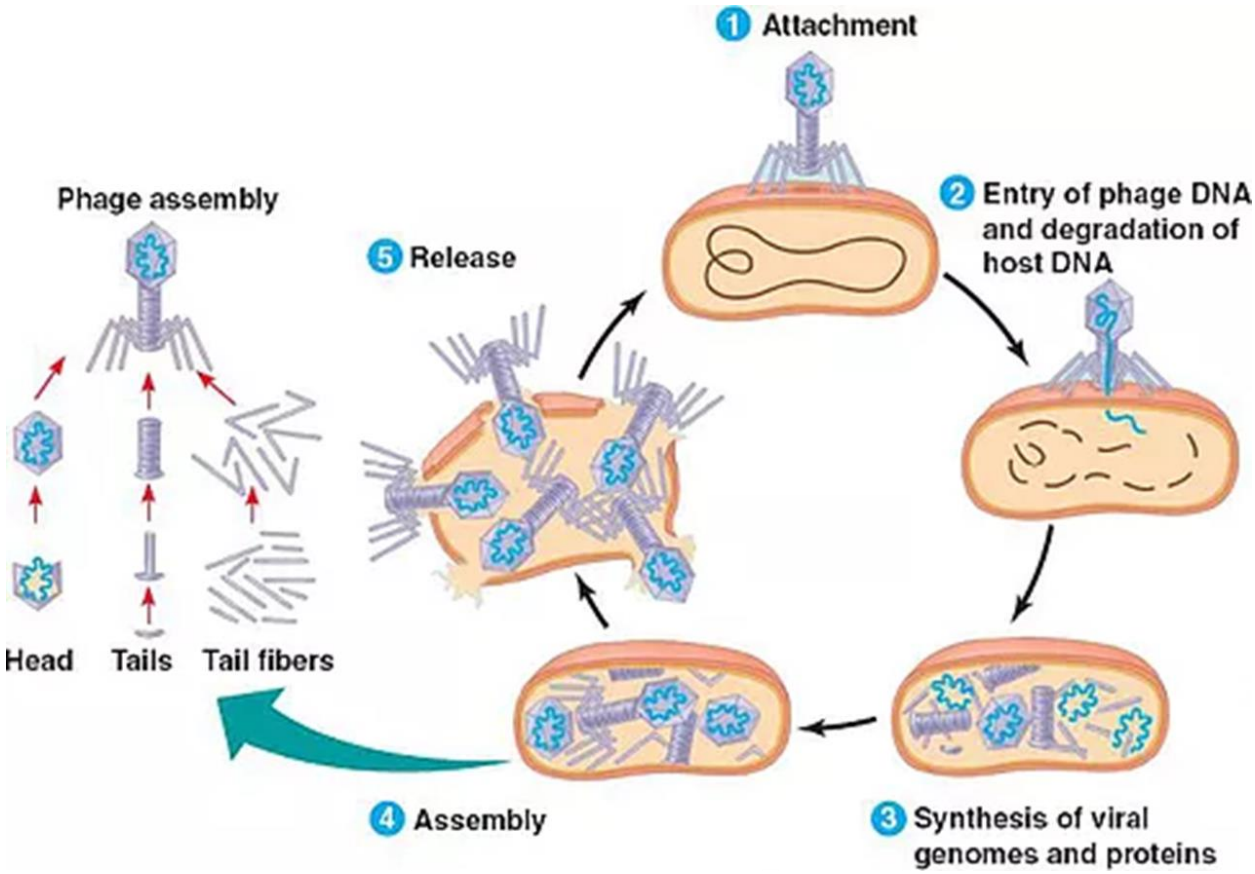
1.1 Phage therapy and its problem

Bacteriophages have drawn much attention due to their potential as an alternative therapy to antibiotics since their discovery. The therapeutic application of phages is referred as to phage therapy and has several advantages over the use of chemical drag [1]. For example, phage has lower toxicity since it mostly consists of mere nucleic acid and protein. Besides, phage therapy has minimal disruption of normal flora due to its narrow host range, which typically covers a few strains of one bacterial species. Furthermore, and most importantly, phage can even treat antibiotic resistant bacteria as the resistance mechanisms of bacteria against phage and antibiotics are substantially distinct. Thus the emergence of antibiotic-resistant bacteria nowadays sparked new interest in this method, even though the enthusiasm for phage therapy once had faded upon the development of antibiotics in the 1940's [2].

Understanding the underlying mechanisms of host-phage interaction would be beneficial to practical application of phage therapy. For example, dosing mixture of several strains which have different host specificities (so-called phage cocktail) is known to be effective to treat a wider range of pathogens [3]–[6]. Identification of genetic variations responsible for the differences in infectivity of those phages should provide insightful information to select a proper combination of phage strains for the cocktail. Sequencing of genes playing a key role for phage infection would allow us to anticipate the efficacy of phage dosage from genomic information. Furthermore, modification of such genes by genome editing may make it possible to artificially alterate or expand host specificity of phage. That is also consider to be useful for phage therapy in practice [7].

However, the genes related to infectivity of phage and phage-resistance of bacteria are

considered to be immensely diverse and so far little understood. Any steps of phage's life cycle (Fig. 1-1) can be targeted by phage-resistant bacteria and phage can acquire counter-adaptation to their resistance mechanism in turn [8]. Through this evolutionary arm-race, bacteria invented diverse phage defense systems and phage developed an assortment of devices to inactivate such barricades [9]. In this study, I tackled this issue through the analysis of two bacteriophages, PP01 and T2.



https://www.brainscape.com/flashcards/virus-structure-amp-lytic-vs-lysogenic-cy-2200902/packs/3787746

Fig. 1-1 Infection process of lytic phage

Phage firstly adsorbs to the surface of the host (1), and then injects its genomic DNA into the host cell (2). Phage proteins and genomic DNA are synthesized (3). The proteins assembled to form mature virus particle and the replicated DNA is packed in the phage capsid (4). The cell wall of bacteria is destroyed by lysis protein to release progeny from the host cell (5).

1.2 PP01 and T2 phage

PP01 phage is a member of the *Escherichia coli* O157:H7-specific T4-like phages, together with AR1 [10], ECML-134 [11], O157-typing phage 7 [12], vB_EcoM_112 [13], and wV7 [14]. It has the potential to be applied for treatment of *E. coli* O157:H7 infection [5]. *E. coli* O157:H7 is an important and well-known pathogen in humans, causing serious diseases such as thrombotic thrombocytopenic purpura and hemolytic-uremic syndrome through contaminated food [15], [16]. Prophylaxis of the food poisoning caused by this pathogen is a big concern of health-care and phage therapy is considered to be a hopeful remedy for this problem [17]. Therefore it should be beneficial to the society to analyze the interaction of *E. coli* O157:H7 and PP01 for the increase of its potential as an antimicrobial agent. Meanwhile, T2 phage is one of the best-characterized T4-like virus and assumed to be closely related to PP01. Nonetheless, it has no infectivity to *E. coli* O157:H7. Therefore the comparison of PP01 and T2 is useful to understand the PP01's mechanism to specifically infect *E. coli* O157:H7. In this study, genetic variation of PP01 and T2 accounted for the difference of their infectivity was explored by whole genome comparison and homologous recombination between those two phages.

1.3 Previous study and its unsolved problems

T4-like virus adsorbs to host cell surface with the tip of the long tail fiber [18], enabling themselves to scan the cell surface for secondary irreversible attachment via the bacterial-lipopolysaccharide (LPS) by the short tail fiber [19] (Fig. 1-2). The long tail fiber of PP01 and T2 consists of five proteins, gp34, gp35, gp36, gp37, and gp38 (Fig. 1-3), whereas their short tail fiber is the single spike-like protein gp12 [20].

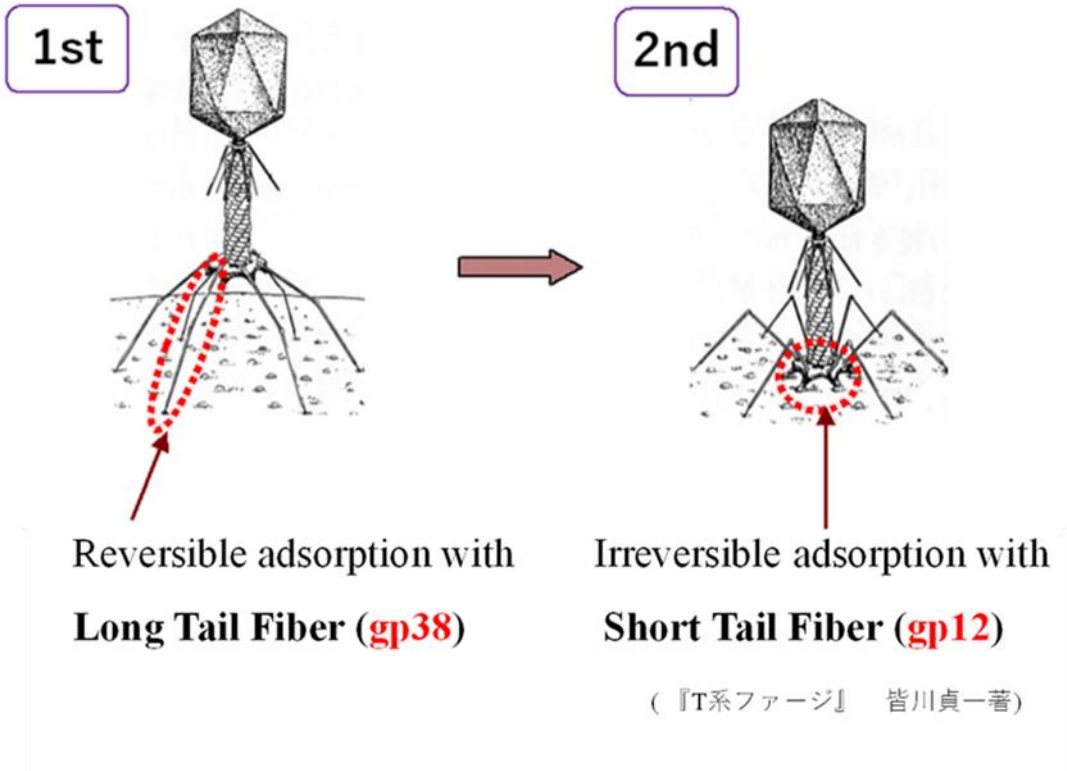


Fig. 1-2 Adsorption to bacterial host surface by PP01 and T2

Those phages firstly attach to LPS or outer membrane protein in the host cell surface via the tip of long tail fiber (gp38) in reversible way. The attachment of the long tail fiber trigger a structural rearrangement of the baseplate, which deploys the six short tail fibers located inside of the baseplate. The short tail fibers irreversibly bind with LPS of the host bacteria.

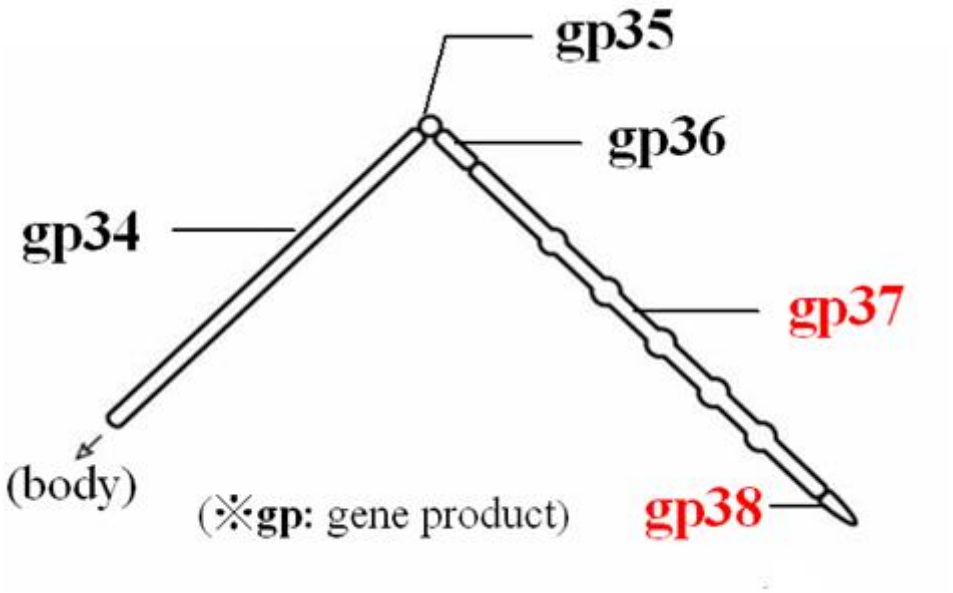


Fig. 1-3 Structure of long tail fiber of PP01 and T2

gp34 and gp37 are the proximal and distal long tail fibers, respectively. gp38 is the tip of distal long tail fiber. gp35 and gp36 form the junction between proximal and distal long tail fibers [20].

The long and short tail fibers of PP01 and T2 were quite varied in their amino acid sequences. Since the adsorption to host cell surface is a primary and essential step for phage infection, those variations were expected to be responsible for the difference of their infectivity. Therefore, in the previous study, genes *37* and *38* of T2, which encode gp37 and gp38 respectively, were replaced with the counterpart of PP01 by homologous recombination, attempting to make T2 able to infect *E. coli* O157:H7 [21]. PP01's genes *37* and *38* were cloned into a plasmid and introduced into *E. coli* C, the strain susceptible to both PP01 and T2. That host was infected with T2 phage and the resulting lysate was screened by infectivity to *E. coli* O157:H7. The isolated recombinant was named T2ppD1 and capable of infecting *E. coli* O157:H7. Thus T2ppD1 elucidated PP01's specificity to *E. coli* O157:H7 relied on its distinctive long tail fiber.

However, infectivity of T2ppD1 to *E. coli* O157:H7 is not comparable with PP01. Firstly, T2ppD1's adsorption to the cell surface of *E. coli* O157:H7 is significantly slower than PP01. It is speculated the short tail fiber of T2ppD1, which is derived from T2, cannot attach with LPS of *E. coli* O157:H7 as efficiently as one of PP01. Since the rate of adsorption to host bacteria is reported to be important parameter for phage fitness [22]–[25], it is of interest to phage therapy whether modification of the short tail fiber can improve the adsorption ability of T2ppD1. Nevertheless, that has been uncertain due to the difficulty to manipulate gene *12*. Secondly, T2ppD1 loses infectivity to *E. coli* O157:H7 once it was grown using *E. coli* C as host. Although T2ppD1 can be serially passaged on *E. coli* O157:H7, after infection to *E. coli* C, the only 10^{-6} of the progeny can infect to *E. coli* O157:H7, whereas PP01 cannot be affected by a host for propagation. This result has great similarity to the phenomenon commonly known as "host-controlled modification", which is involved with methylation pattern of phage genome [26], [27].

However, the exact mechanism behind incomplete infectivity of T2ppD1 still remains to be revealed.

In this dissertation, using CRISPR/Cas9 system, both long and short tail fibers of T2 was replaced with the counterpart of PP01 to endow adsorption ability comparable to PP01 with the recombinant phage (chapter 2). Besides, whole genome comparison was conducted to seek PP01's genes essential for infection to *E. coli* O157:H7 other than the gene of the long tail fiber. The highly variable genes between PP01 and T2 were chosen as candidates and further narrowed down by quick experimental screening based on CRISPR/Cas9 (chapter 3). As a result of that, gene *motB* was picked up as the most possible candidate. Its role for infection to *E. coli* O157:H7 was analyzed by knock-out of *motB* from PP01 and exchange of both long tail fiber and *motB* of T2 with the respective genes of PP01 aiming at enabling the T2 recombinant to infect *E. coli* O157:H7 (chapter 4).

CHAPTER 2 Modification of the long and short tail fibers of T2 phage by CRISPR/Cas9

2.1 Introduction

In this chapter, it was firstly attempted to apply CRISPR/Cas9 system for genome editing of T2 phage. Then the long and short tail fibers of T2 were replaced with the counterpart of PP01 using CRISPR/Cas9 to gauge whether adsorption ability to *E. coli* O157:H7 can be improved by alteration of the short tail fiber.

An efficient genome editing method for T2 phage was necessary to manipulate gene *12*. Although an elaborated genome editing method such as insertion-substitution vector system is already available for T4 phage, the system requires use of the special mutant (T4 38^{am}, 51^{am} $denA^{-}$ $denB^{-}$) [28]. Those mutations are laborious to introduced into T2 and also undesirable to evaluate adsorption ability and infectivity of the resultant recombinant. Meanwhile, several works have recently reported efficient phage genome editing by the CRISPR/Cas system [29]–[33]. In particular, Tao et al. showed the successful genome editing of the T4 phage genome, with the potential to be applied likewise to other T4-like phages [33].

2.2 Material and method

2.2.1 Bacterial and phage strains

Four strains of *E. coli* were used in this study: *E. coli* DH10B (Promega Co., Ltd, USA), *E. coli* DH10BΔgalU (*galU* knock-out mutant of *E. coli* DH10B), *E. coli* C [34] and *E. coli* O157:H7 (ATCC43888). *E. coli* DH10B are DH10BΔgalU sensitive to T2 and resistant to PP01. Conversely, *E. coli* O157:H7 is susceptible to PP01 and unsusceptible

to T2. Meanwhile, *E. coli* C is sensitive to both phage strain (Table 2-1). Host bacteria were grown at 37ºC in LB media. 50 μg/ml ampicillin and/or 30 μg/ml chloramphenicol were added to the media for plasmid selection. *E. coli* DH10BΔgalU was constructed by replacing its *galU* gene with a kanamycin resistance gene by lambda recombination according to the protocol described in Wilson *et al*. [35]. Lambda recombination system was provided by pKD46 plasmid extracted from *E. coli* MG1655/pKD46 (CGSG no. 7669). The kanamycin resistance gene of Tn5 was amplified using the primers shown in Table 2-2.

PP01 is *E. coli* O157:H7 specific phage isolated from pig feces [5] and T2 is one of the classic phages, which has no infectivity to *E. coli* O157:H7. Three T2 mutants, $T2_{PP01g37\text{-}38}$, $T2_{PP01g12}$, and $T2_{PP01g37\text{-}38\text{-}12}$ were constructed as described in following sections (Table 2-1). The lysates of T2, T2-derived mutants and PP01 were propagated by plate lysate method using *E. coli* C as host. One hundred μl of the phage lysate of around 10^5~10^6 PFU(Plaque-Forming Unit)/ml was mixed with 100 μl of the overnight host culture in 3 ml of 0.5% top agar, plated on LB agar plate, and incubated at 37°C for overnight. Then 5 ml of salt magnesium (SM) buffer (100 mM NaCl, 8 mM $MgSO_4$, 50 mM Tris-HCl (pH 7.5), 0.01% gelatin) was added to the plate. The top agar was scraped off and centrifuged (5000 g, 15 min, 4°C) to recover the phage lysate as supernatant. The lysate was purified by polyethylene glycol (PEG) precipitation.

Table 2-1 Bacteria and phage strains

Strain name	Description
Bacteria	
E. coli K12 DH10B	Susceptible to T2
E. coli K12 DH10BΔgalU	The *galU* knock-out DH10B mutant
E. coli C	Susceptible to both PP01 and T2
E. coli O157:H7 (ATCC43888)	Susceptible to PP01
Bacteriophage	
T2	Wild-type
$T2_{PP01g37\text{-}38}$	T2 recombinant with PP01's long tail fiber (gene *37* and *38*)
$T2_{PP01g12}$	T2 recombinant with PP01's short tail fiber (gene *12*)
$T2_{PP01g37\text{-}38\text{-}12}$	T2 recombinant with PP01's long and short tail fibers
PP01	O157:H7 phage

Table 2-2 Primers for construction of DH10BΔgalU

Primer name	Sequence
Tn5-DH10B-galU-F	ATACAGAAATATGAACACGTTCAAAACACGAACAGTCCAGG AGAATTTAATATGGACAGCAAGCGAACCG
Tn5-DH10B-galU-R	CGTCGATTGCTCAACGCCGTTTCGTGGATAACACCGATACG GATGTCAGAAGAACTCGTCAAGAAG

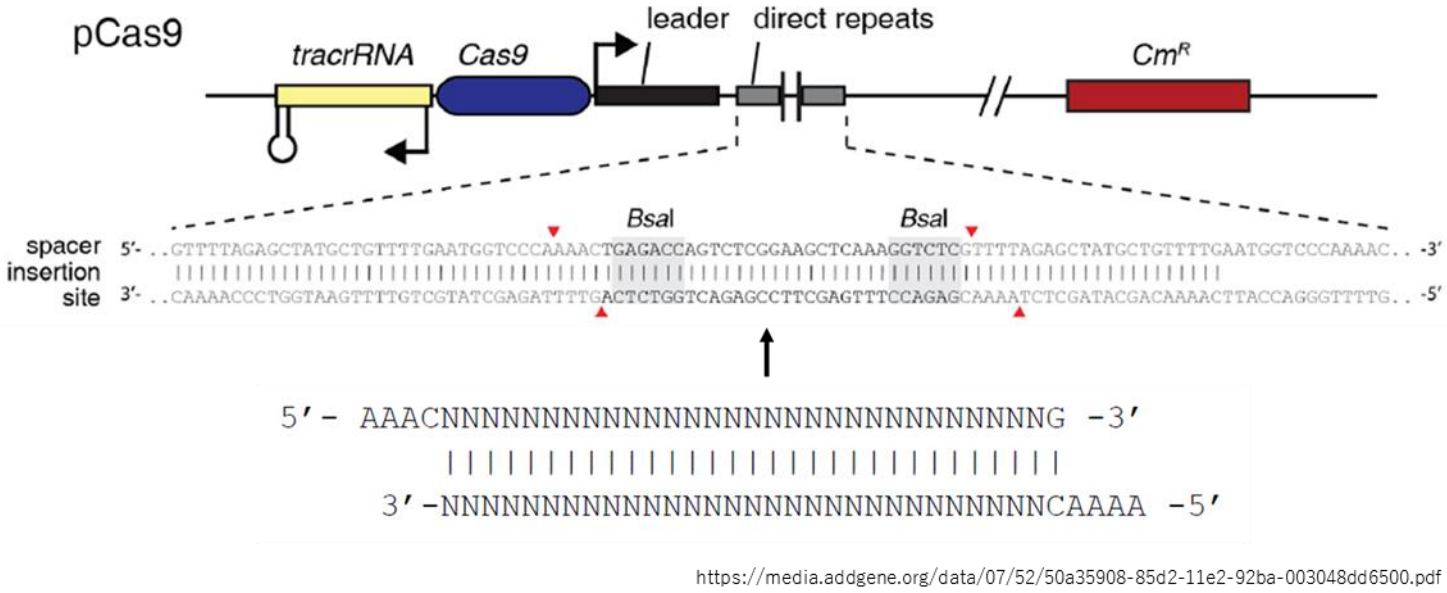
2.2.2 CRISPR/Cas9 system

In brief, CRISPR/Cas9 is a DNA endonuclease whose specificity is determined by a sequence of a short RNA called crRNA. crRNA forms protein-RNA complex with Cas9 endonuclease and the complex selectively attacks DNA sequence complementary with a certain part of crRNA. That part of crRNA and its complimentary target DNA is referred to as spacer sequence and protospacer sequence, respectively. Thereby, any DNA sequence can be basically targeted by modifying spacer sequence [29].

The plasmid pCas9, which encodes CRISPR/Cas9 derived from *Streptocossus. pyogenes* SF370 was used in my study. It was kindly provided by Luciano Marraffini (Addgene plasmid no. 42876) [36]. The vector map of pCas9 was shown in Fig. 2-1. The protocol to modify its spacer was described in the supplemental document of pCas9 in Addgene website. (https://media.addgene.org/data/07/52/50a35908-85d2-11e2-92ba-003048dd6500.pdf). Shortly, Spacer sequences were synthesized as two single-strand oligonucleotides, which were homologous with a certain part of T2 phage genome. In order to insert those oligos to the cloning site of pCas9, AAAC and G were put at 5' and 3' end of the plus strand oligo respectively, and AAAAC was added into 5' end of the minus strand oligo (Fig. 2-1). 5' ends of those two oligo nucleotides were phosphorylated by T4 polynucleotide kinase (New England Biolab). Those oligos were annealed by heating at 95℃ for 5 min and slowly cooled down from 95°C to 25°C for 1 hour. pCas9 was digested by restriction enzyme BsaI (New England Biolab) and ligated with the annealed oligos by T4 ligase (New England Biolab) to transform *E. coli* K12 DH10B competent cell.

The gene *denA*, gene *38* and gene *12* were chosen as the targets of CRISPR/Cas9. The sequences of the spacers and the names of the plasmids constructed by cloning those into

pCas9 are shown in Table 2-3. They were designed based on the genomic sequences of PP01 and T2 (available in DDBJ under accession number LC348379 and LC348380). The gene *denA* encodes exonuclease to degrade host genomic DNA during infection [20] and randomly picked from non-essential genes for a proof-of-concept of T2 genome editing by CRISPR/Cas9. The gene *38* and gene *12* are encoding the long and short tail fibers, respectively, and targeted to replace those genes with the corresponding parts of PP01.



https://media.addgene.org/data/07/52/50a35908-85d2-11e2-92ba-003048dd6500.pdf

Fig. 2-1 Vector map of pCas9 and the cloning site of spacer sequence

The vector map of pCas9 is shown at top. It consists of the regions coding trans-crRNA, Cas9, crRNA (composed of the leader sequence and the spacer sequence flanked by two direct repeats), and a chloramphenicol resistance gene. The spacer sequence has BsaI restriction site. Thereby synthesized oligo DNAs (typically 20~30 bp) designed as shown at bottom can be introduced into that region.

Table 2-3 Plasmids and host bacteria for the recombination experiments

Plasmid name		
CRISPR/Cas9	Target	Spacer sequence
pCas9-g38	gene *38* of T2	CGGTGTGACTGTTTATGGCCGAGGAGGTAA
pCas9-g12	gene *12* of T2	AATGTATATCTGGAATTTCGTTGTGATGCT
pCas9-denA	*denA* of T2	TTAAAAAATATTTTTAAGGAGTGGGCCGCA
Editing template		
pK-denA	For knock-out of *denA* of T2	
pPg12Ca	For replacing the short tail fiber (gene *12*) of T2 with one of PP01	
pPg3738	For replacing the long tail fiber (gene *37* and *38*) of T2 with one of PP01	

Combinations of the hosts and the plasmids for the recombination experiment			
Modified gene	Host	Recombination	Selection
denA	DH10BΔgalU	pCas9-denA & pK-denA	pCas9-denA
gene *12*	DH10BΔgalU	pCas9-g12 & pPg12Ca	pCas9-g12
gene *37-38*	C	pCas9-g38 & pPg3738	pCas9-g38

2.2.4 Editing templates

An editing template is a DNA fragment homologous with a part of phage genome and carrying desired mutations. The plasmids carrying editing templates to modify gene *denA*, gene *12* and gene *37-38* were named pK-denA, pPg12Ca, and pPg3738, respectively (Table 2-3). The schematic structures of the editing templates used in this study are represented in Fig. 2-2. The primers used to construct the editing templates are displayed in Table 2-4. PCR products amplified from T2 or PP01 genomic DNA was fused using In-Fusion™ (Clontech Co., Ltd, USA) and cloned into pUC118.

The editing template for knock-out of *denA* was designed to introduce stop codons nearby its start codon. Three sequential stop codons were used to prevent run-on translation and ensure loss of the protospacer in T2 genome by homologous recombination. PCR was performed using T2 genomic DNA as template and primer pairs denA-k-up-F/denA-k-up-R and denA-k-dw-F/denA-k-dw-R to amplify upper and lower half parts of the editing template, respectively. denA-k-up-R and denA-k-dw-F had 15 bp homology regions for fusion, which included three ochre codons (TAA-TAA-TAA).

For modification of the short tail fiber, the editing template was designed to swap only C-terminal part of gp12 of PP01 and T2, since the first attempt to recombine entire gene *12* was unsuccessful. According to the structural characterization of T4 gp12 [37], it seems to attach with the base plate by the N-terminal part and bind to LPS of host bacteria via the tip of the C-terminal part (which includes the residues 397-517). I presumed that the structures of gp12 of PP01 and T2 were similar to the one of T4. Thereby the part of PP01 gp12 (residues 390-507) was assumed to be the tip of PP01's short tail fiber based on their sequence alignment (Fig. 2-3). This part of PP01 genome was amplified using primer pair PP01-g12-Ca-F/PP01-g12-Ca-R. The parts of T2 genome corresponding to

the flanking regions of PP01's gp12 C-terminal part were amplified using primer pairs T2-g12-Ca-up-F/T2-g12-Ca-up-R and T2-g12-Ca-dw-F/T2-g12-Ca-dw-R.

For recombination of gene *37* and *38*, the same editing template constructed in our previous work was used [21]. Although, the gene *37* and *38* of PP01 aren't quite similar to one of T2, their ~100 bp flanking regions have homology with the corresponding part of T2 genome.

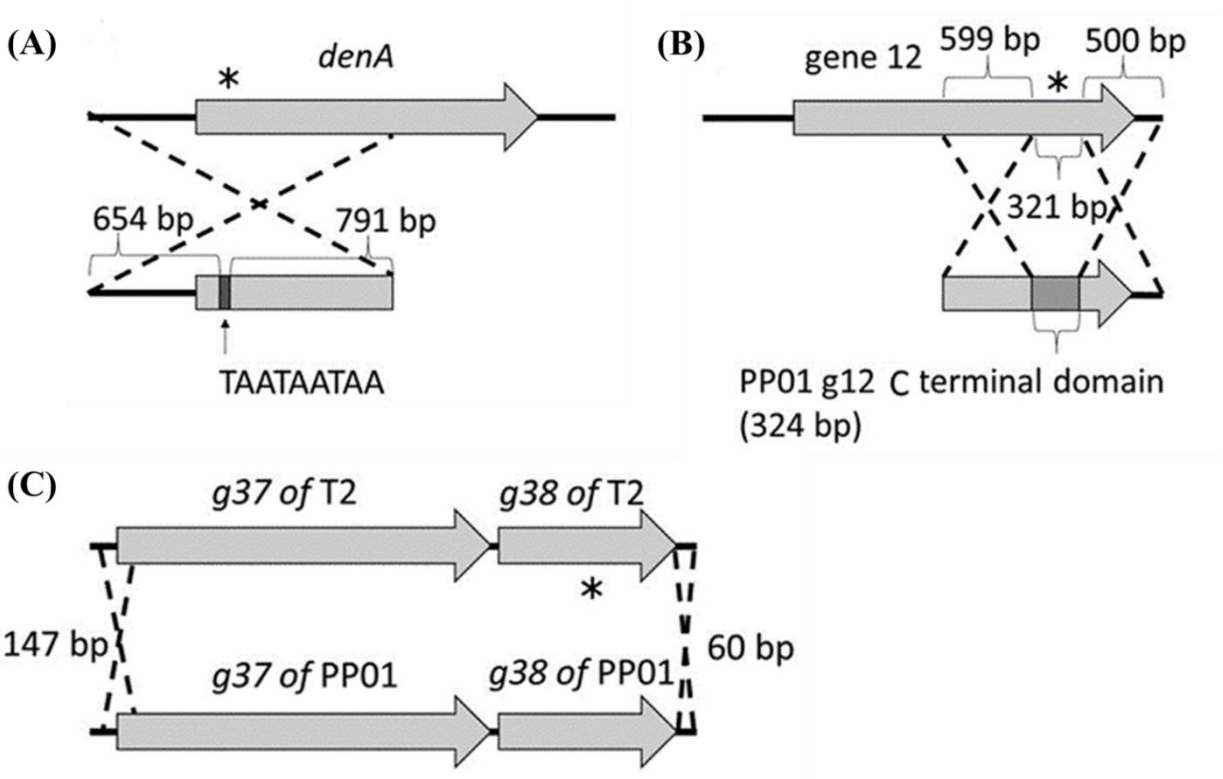


Fig 2-2 Structure of the editing templates

The structure of the editing templates to modify *denA* (A), gene *12* (B) and gene *37-38* (C) are displayed. The arrows indicate the ORFs of PP01 or T2. The dashed lines represent the regions homologous between the editing templates and the phage genome. The numbers indicate the length of those regions. The asterisks show the approximate positions of the proto-spacer sequences.

Table 2-4 Sequences of the primers for construction of the editing templates

Primer name	Plasmid	Sequence
denA-k-up-F	pK-denA	CGGTACCCGGGGATCCTTGATACCAAAGACCCATCTTC
denA-k-up-R	pK-denA	GTGGGCTAATAATAACATTTTATTATGAAAGAAATTGCAACAG
denA-k-dw-F	pK-denA	TTATTATTAGCCCACTCCTTAAAAATATTTTTTAAAACTCATCATAAC
denA-k-dw-R	pK-denA	CGACTCTAGAGGATCCGAAATTAACGCAGACGAAGATG
T2-g12-Ca-up-F	pPg12Ca	ATCCTCTAGAGTCGACTTAACTCGAATAATGCTTC
T2-g12-Ca-up-R	pPg12Ca	ACCGCCGGTGCCCGCGCCACGAACAAAAAGACC
PP01-g12-Ca-F	pPg12Ca	GCGGGCACCGGCGGTCATATTTTAAATC
PP01-g12-Ca-R	pPg12Ca	ATAACCAATTAATCCTTCGCGATTTAAC
T2-g12-Ca-dw-F	pPg12Ca	GGATTAATTGGTTATGAAACACGTCCATGGAAC
T2-g12-Ca-dw-R	pPg12Ca	ATGCCTGCAGGTCGACCAGTAAATCATTTCTGATTG

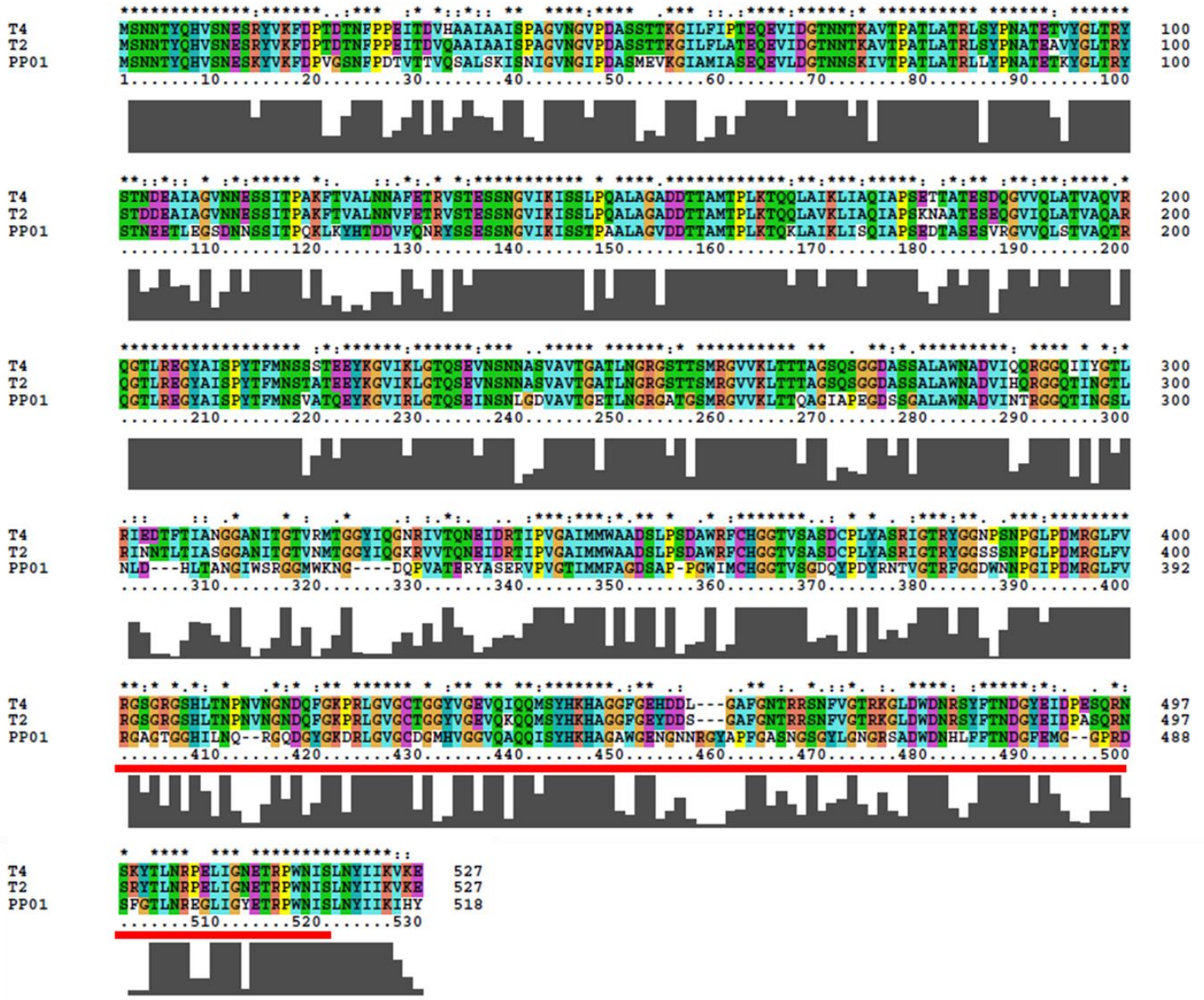


Fig. 2-3 Multiple alignment of gp12 of T4, PP01 and T2

The hypothetical LPS attaching part of gp12 of PP01 and T2, which was swapped by homologous recombination, is underlined by the red line.

2.2.4 EOP test

Efficiency-of-plating (EOP) was measured to confirm the inhibition of T2 phage infection by CRISPR/Cas9. EOP was calculated by dividing PFU number estimated from plating on the bacterial host containing CRISPR/Cas9 by one measured on the same host without CRISPR/Cas9. EOP of T2 phage was measured using *E. coli* DH10B, *E. coli* DH10BΔgalU, and *E. coli* C carrying pCas9-derived plasmids shown in Table 2-3.

2.2.5 Homologous recombination and screening of recombinant phage

The host bacteria were transformed by the plasmids carrying an editing template and CRISPR/Cas9 and then infected with T2 phage. The spacer sequence was designed in such a way that it could only recognize wild-type phages and unable to restrict recombinant phages, as conducted in previous studies [29]–[33]. Thus, CRISPR/Cas9 select phages which occurred homologous recombination with the editing template during infection (Fig. 2-4). The progeny were further selected by plating on the hosts with only CRISPR/Cas9. The combination of the host and the plasmids for recombination and selection were summarized in Table 2-3.

In detail, 50 μl of O/N culture of the host strain carrying an editing template and CRISPR/Cas9 were inoculated into 5 ml of LB medium and cultured until reaching early growth phase (OD_{600} ~ 0.2). Then the phage lysate of T2 was added to the culture with MOI equal to ~0.01 and incubated for over-night for homologous recombination. The culture was centrifuged and the phage lysate was recovered as supernatant. The lysate was filtered with a 0.45 μm filter for sterilization and plated on the same host strain carrying only CRISPR/Cas9 for selection. Ten plaques were tested to confirm recombination by plaque direct PCR using the Mighty amp PCR kit (Takara Co., Ltd,

Japan). The primers used for screening are reported in Table 2-5.

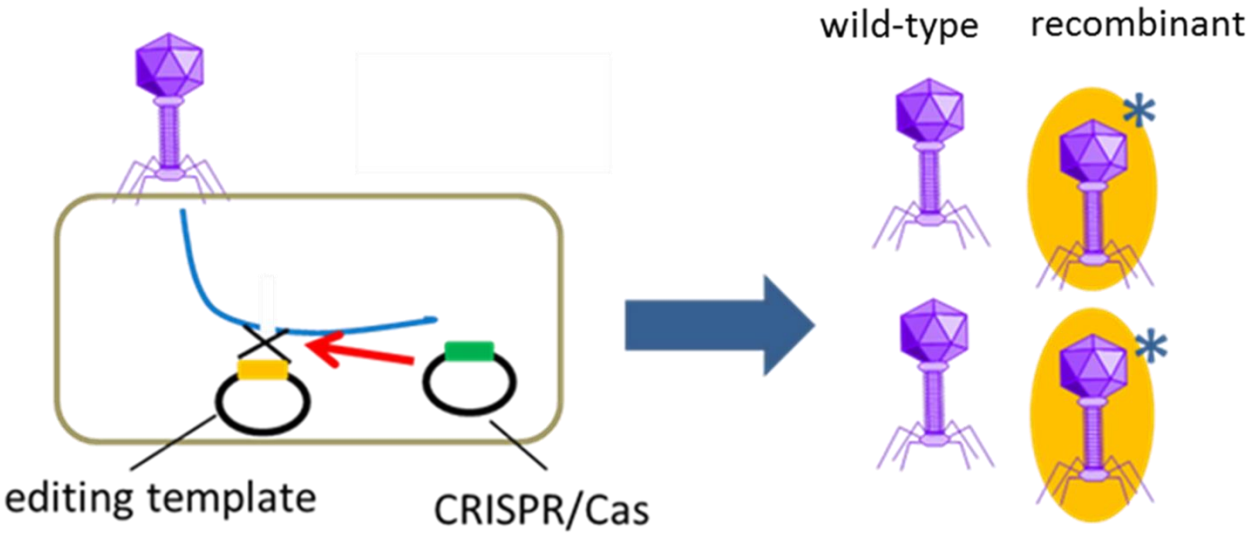


Fig. 2-4 Principle of genome editing by CRISPR/Cas9

CRISPR/Cas9 cleaves the targeted site of phage genome (protospaer). The editing template is homologous with the region flanking to the protospacer but doesn't include the protospacer ifself. Thus the recombinant phages which incorporate the editing template into its genome by homologous recombination can escape from CRISPR/Cas9 and become dominant in the progeny.

Table 2-5 Sequences of the primers for screening of recombinants

Primer name	Target gene	Sequence
denA-seq-f	*denA* (fw)	GGCATACATTGCTTTAAGATC
denA-seq-r	*denA* (rv)	AAGGAGTGGGCTAATAATAAC
g38-seq-f	*38* (fw)	TGGGCCGCTGGTGCAATTG
g38-seq-r	*38* (rv)	TGCATAATACCCTGTACCAGTATTACC
g12-seq-f	*12* (fw)	ATGAGTAATAATACATATCAA
g12-seq-r	*12* (rv)	TAATGGCAGTGGATATTTAGGCAA

2.2.6 Adsorption assay

200 μl of O/N culture of *E. coli* O157:H7 was inoculated into 2 ml of LB medium and incubated for 1 hour. The culture was diluted to OD_{600} 0.1. The phage lysate of T2, $T2_{PP01g12}$, $T2_{PP01g37\text{-}38}$ or $T2_{PP01g37\text{-}38\text{-}12}$, or PP01 was added to 4.5 ml of the diluted culture with MOI 0.01. The mixture of host and phage was incubated at 37°C under agitation (120 rpm) and sampled after 0 and 15 minutes. The sample was diluted 100 fold by SM buffer immediately after collection to stop adsorption of phage to host. 1 ml of this sample was taken and the cell of *E. coli* O157:H7 and the adsorbed phage were removed by centrifugation (21,880 g, 1 min). In the end, the number of the phage particles remaining in the supernatant was enumerated by plaque assay on *E. coli* C strain.

2.2.7 Comparison of gp12 sequence among *E. coli* O157:H7 specific phages

The accession numbers of genome data of T2 phage and *E. coli* O157:H7 specific phages (AR1, ECML-134, O157-typing phage 7, PP01, vB_EcoM_112, wV7) are LC348380, NC_027983, NC_025449, LC348379, KP869105, NC_024125, and NC_019505, respectively. The amino acid sequences of gp12 were extracted from their genomic sequences and compared by EMBOSS needle program [38]. The heat map was drawn by heatmap.2 function of R [39].

2.3 Results

2.3.1 Inhibition of T2 phage infection by CRISPR/Cas9

In order to exploit CRISPR/Cas9 for genome editing of T2, it was gauged whether CRISPR/Cas9 could work against this phage to inhibit its infection. Besides, the effect of T2 genome modification on the efficiency of CRISPR/Cas9 was also examined.

The genomic DNAs of T4-like phages including PP01 and T2 commonly undergo 5-hydroxymethylation and glucosylation of cytosines to avoid host restriction enzymes [40]. That modification was expected to also block CRISPR/Cas system. In fact, it was shown that T4 phages carrying DNA containing glucosyl-hydroxylmethylated cytosines (ghmC DNA) were more resistant against CRISPR/Cas9 compared to the T4 mutants carrying unmodified DNA (T4C) depending on spacer sequences [33]. The glucosylation of T2's cytosine is dependent on *galU* of host bacteria and *α-gt* of T2 phage. Thus the genomic DNA of T2 propagated on the galU^{-} host is composed of hydroxymethylate cytosine (hmC DNA) and more susceptible to restriction system than ghmC DNA [41]. Therefore, DH10BΔgalU carrying CRISPR/Cas9 was expected to show stronger inhibition of T2 phage infection.

The results of EOP test are shown in Fig 2-5. EOP was here defined as the ratio of PFU measured on the host carrying CRISPR/Cas9 to PFU measured on the host without CRISPR/Cas9, which means the reduction of EOP represented the effect of CRISPR/Cas9 to inhibit phage infection. In all tested conditions, T2 phage infection was hampered by CRISPR/Cas9. Although the efficiency was poor in most cases (~10^{-1}), *E. coli* DH10BD*galU* carrying pCas9g12 displayed significantly stronger inhibition of infection (~10^{-4}) (Fig. 2-5). Thus it was shown that CRISPR/Cas9 could successfully work against T2 phage and also implied T2 phages carrying hmC DNA could be more susceptible to

CRISPR/Cas9, although it depended on the spacer sequences, similarly to the report of Tao et al. [33].

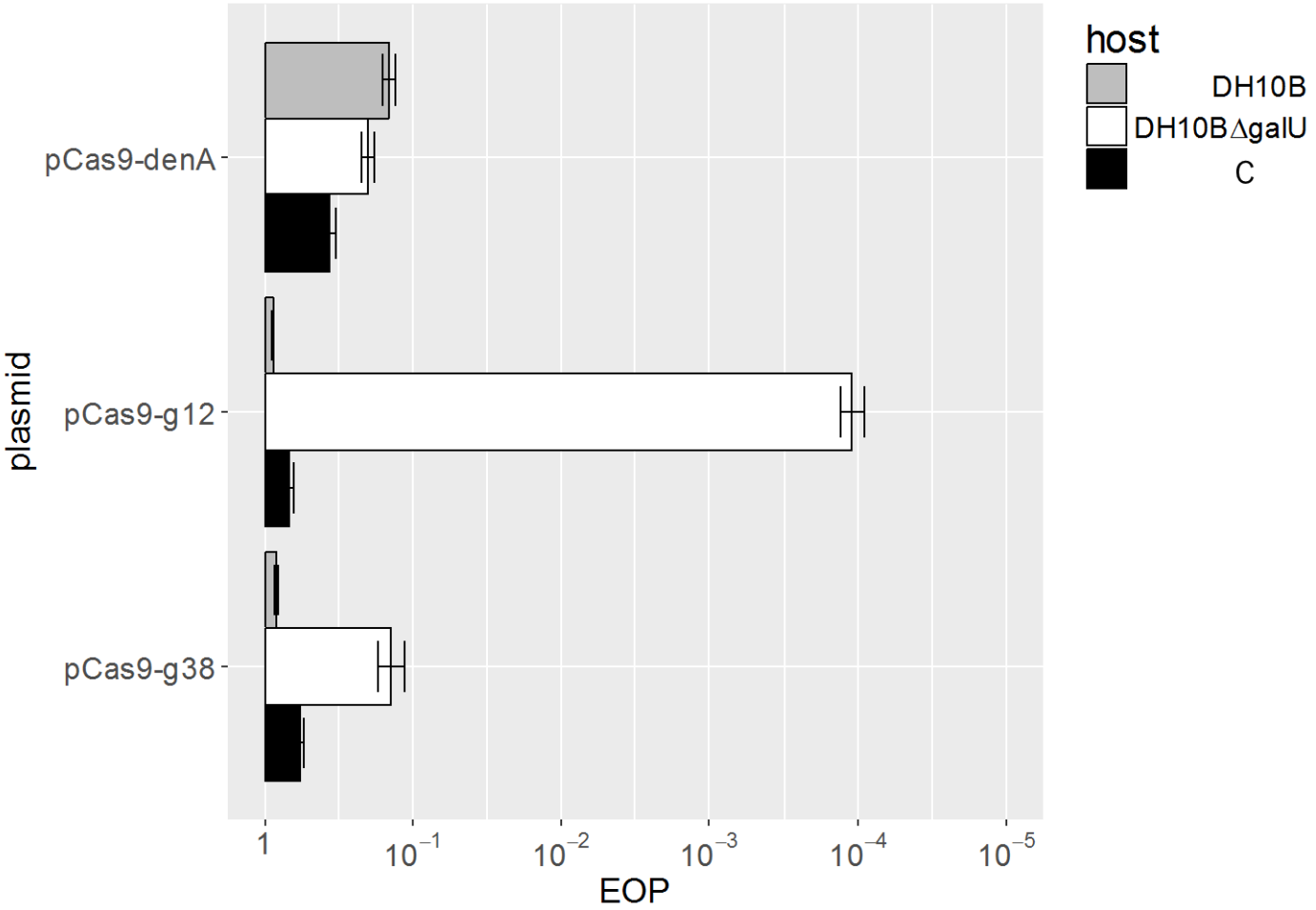


Fig. 2-5 Efficacy of CRISPR/Cas9 against T2 phage

EOP was defined as the ratio of phage titer estimated on the host (*E. coli* DH10B, DH10BΔgalU, and C) carrying CRISPR/Cas9 to the same host without CRISPR/Cas9. Thus the reduction of EOP represents the extent of inhibition of infection by CRISPR/Cas9. The data represent means ± standard deviations of three technical replicates.

2.3.2 Genome editing of PP01 and T2 by CRISPR/Cas9

The modification of gene *denA, 12* and *37-38* of T2 were conducted using the plasmids and hosts shown in Table 2-3. The screening of ten plaques was sufficient to obtain recombinants for modification of *denA* and gene *12*, whereas recombinant of gene *37* and *38* couldn't be isolated (Table 2-6). In order to address this problem, the resultant lysate of recombination experiment of gene *37* and *38* was applied for the same experiment again. It was grown on *E. coli* C/pCas9-g38 and pPg3738 and plated on *E. coli* C/pCas9-g38 once more. This resulted in all plaques bearing recombinant of gene *37* and *38* (Table 2-6). It seems that since the recombination of gene *37* and *38* was infrequent and the selection by pCas9-g38 was inefficient, the ratio of recombinant phage was too small to isolate upon screening of 10 plaques. However, the recombinants seemed to prevail upon one more round of recombination and selection.

Table 2-6 Result of the recombination experiments

Modified gene	Plaques(Recombinant/tested)
denA	10/10
gene 12	10/10
gene 37-38	0/10
gene 37-38 (2nd round)	10/10

2.3.3 Characterization of $T2_{PP01g37-38}$ and $T2_{PP01g37-38-12}$

The recombinant of gene *37-38* and gene *12* of T2 phage was named $T2_{PP01g37-38}$ and $T2_{PP01g12}$, respectively. Then gene *37-38* of $T2_{PP01g12}$ was modified in the same way to yield $T2_{PP01g37-38-12,}$ the double mutant of long and short tail fibers, to see whether the modification of the short tail fiber can further improve adsorption to *E. coli* O157:H7.

The results of adsorption assay of T2, $T2_{PP01g12}$, $T2_{PP01g37-38}$, $T2_{PP01g37-38-12}$, and PP01

are shown in Fig. 2-6. T2 phage (negative control) couldn't adsorb to *E. coli* O157:H7 and neither could $T2_{PP01g12}$ as expected. $T2_{PP01g37\text{-}38}$ could adsorb to *E. coli* O157:H7 but its adsorption rate was lower than one of PP01, as previously described [21]. Meanwhile, $T2_{PP01g37\text{-}38\text{-}12}$ displayed higher adsorption ability and it had no significant difference from one of PP01. Thus it was shown that short tail fiber of PP01 was responsible for the higher rate of adsorption to *E. coli* O157:H7.

The infectivity of $T2_{PP01g37\text{-}38}$ and $T2_{PP01\text{-}g37\text{-}38\text{-}12}$ to *E. coli* O157:H7 was also evaluated. They could infect *E. coli* C and couldn't infect *E. coli* DH10B similarly to PP01. Thus their PFU number could be titrated using *E. coli* C as host. Around 10^{-6} of the total population of those recombinants formed plaques on *E. coli* O157:H7 (Fig. 2-7). Those plaques could be isolated and grown on *E. coli* O157:H7. However, once this lysate was propagated on *E. coli* C, the progeny lost infectivity against *E. coli* O157:H7 (data not shown). The traits of $T2_{PP01g37\text{-}38}$ and $T2_{PP01g37\text{-}38\text{-}12}$ were consistent with one of T2ppD1 constructed in the previous study [21].

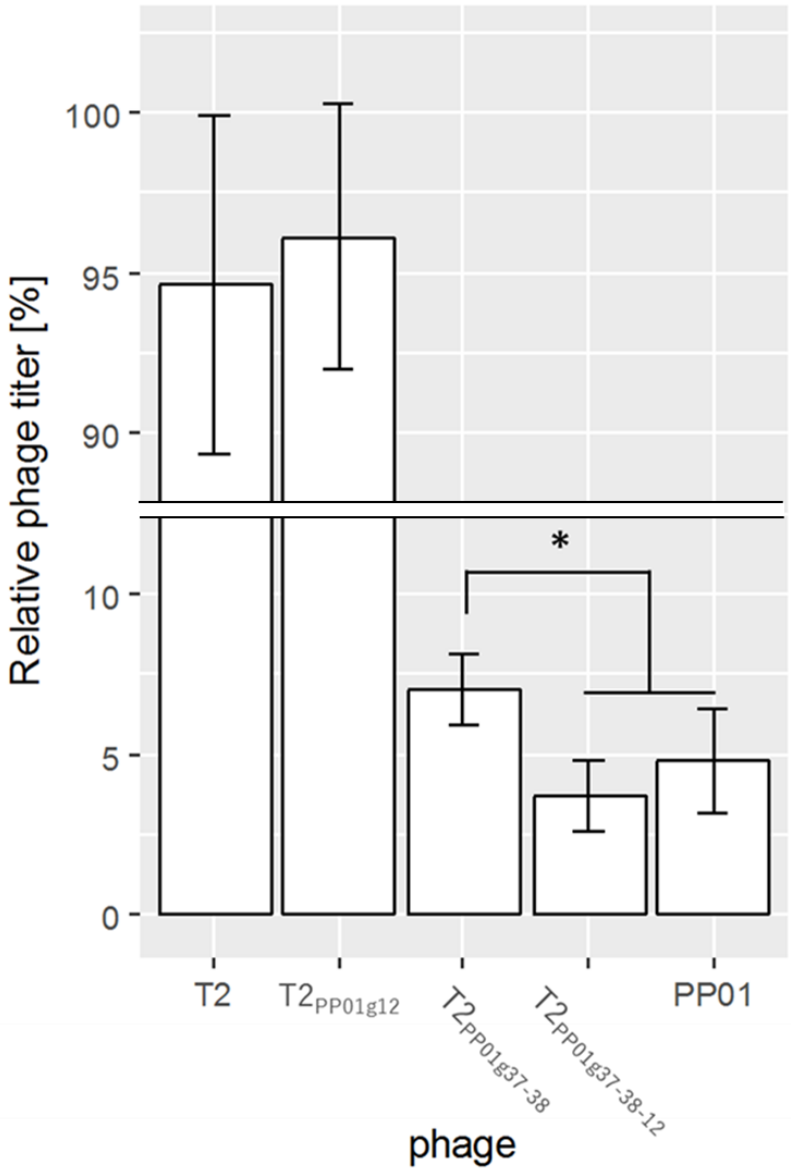


Fig. 2-6 Adsorption assay

The results of adsorption assay of T2, $T2_{PP01g12}$, $T2_{PP01g37\text{-}38}$, $T2_{PP01g37\text{-}38\text{-}12}$ and PP01 on *E. coli* O157:H7 were displayed. The titer of unadsorbed phage at t = 15 min was divided by one at t = 0 min to calculate the relative phage titers. Thus a lower bar represents higher adsorption rate. The data represent means ± standard deviations of three biological replicates. The asterisk represents statistical significance (unpaired Student's t-test, $p < 0.05$).

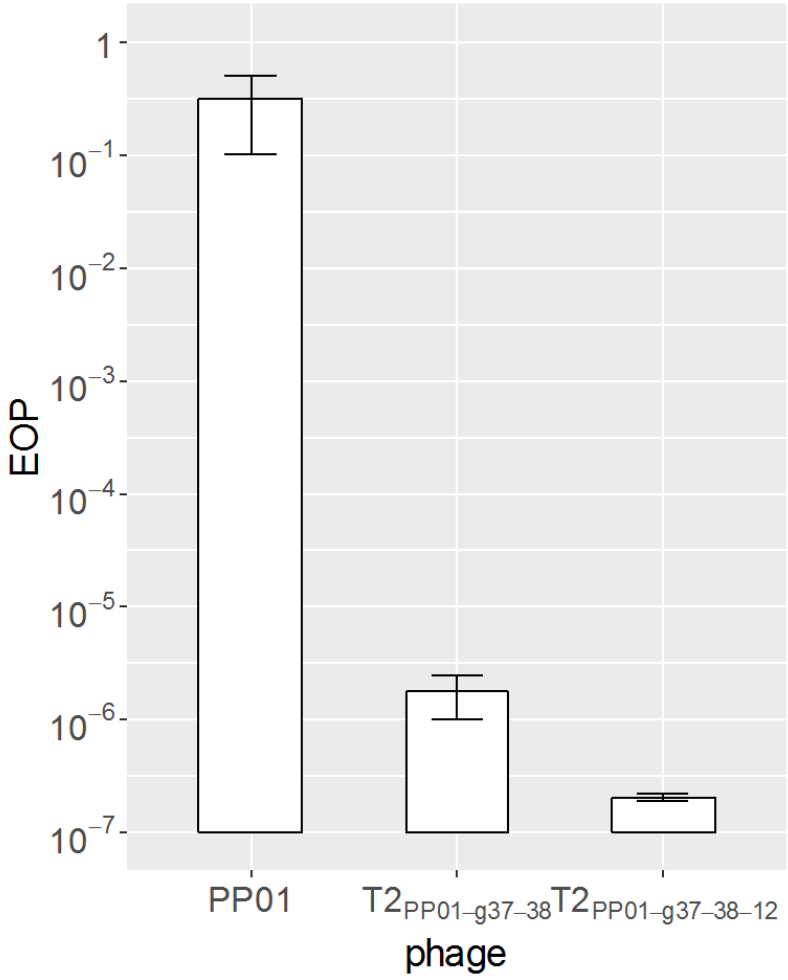


Fig. 2-7 EOP test of PP01 (positive control), T2$_{PP01\text{-}g37\text{-}38}$, and T2$_{PP01\text{-}g37\text{-}38\text{-}12}$

EOP was defined as the plaque count on *E. coli* O157:H7 divided by the plaque count on *E. coli* C. The data represent means±standard deviations of three technical replicates.

2.4 Discussion

In this chapter, it was shown that CRISPR/Cas9 system could be applied to genome editing of T2 phage in a simple and efficient way, as it is implied to be promising for T4-like phage [33]. An efficient method of genome editing for phage is crucial for the analysis of phage biology. Although the efficacy of CRISPR/Cas9 against T2 phage was quite low, recombinant phages could be isolated after one or two sequential rounds of recombination and selection process. This success would encourage the application of CRISPR/Cas9 to other T4-like phages. It should be also noted that hmC DNA could be more susceptible to CRISPR/Cas9 than ghmC DNA. Even if CRISPR/Cas might not work efficiently enough for a particular kind of T4-like phage, preparing phages with hmC DNA might be helpful for that problem. ghmC DNA is a common in T4-like phage [40] but constructing mutant without that modification might be unfeasible. Knock-out of gene *42* (gene of cytosine hydroxylmethylase), which is necessary to obtain cytosine unmodified mutant, requires further mutations to compensate metabolic change caused by loss of gene *42* [42]. Thus preparing phage with hmC DNA, which can be achieved by knock-out glucosyl transferase gene of phage or *galU* of the host, might be more easily applied to other T4-like phages.

In addition, $T2_{PP01\text{-}g37\text{-}38\text{-}12}$, the recombinant of T2 phage whose long tail fiber and tip of the short tail fiber were replaced with the counterpart of PP01, implied the significance of the short tail fiber to adsorption ability to *E. coli* O157:H7, which is known to affect fitness of phage [22]–[25]. Importance of short tail fiber for efficient adsorption might be also true to the other T4-like *E. coli* O157:H7 specific phages, considering the sequences of gp12 are highly conserved among them (Fig. 2-8). This insight will be helpful for construction of artificial *E. coli* O157:H7 phage in the feature.

Meanwhile, it remained to be revealed why T2 phage could not infect to *E. coli* O157:H7. The result apparently showed T2's inability in infecting *E. coli* O157:H7 is not accounted only for adsorption defects due to the variation of the long and tail fiber. This host seemed to inhibit some steps of the infection process of T2 following the attachment to the host surface (DNA entry, genome replication, and so on) (Fig. 1-1). In the next two chapters, this point will be further analyzed.

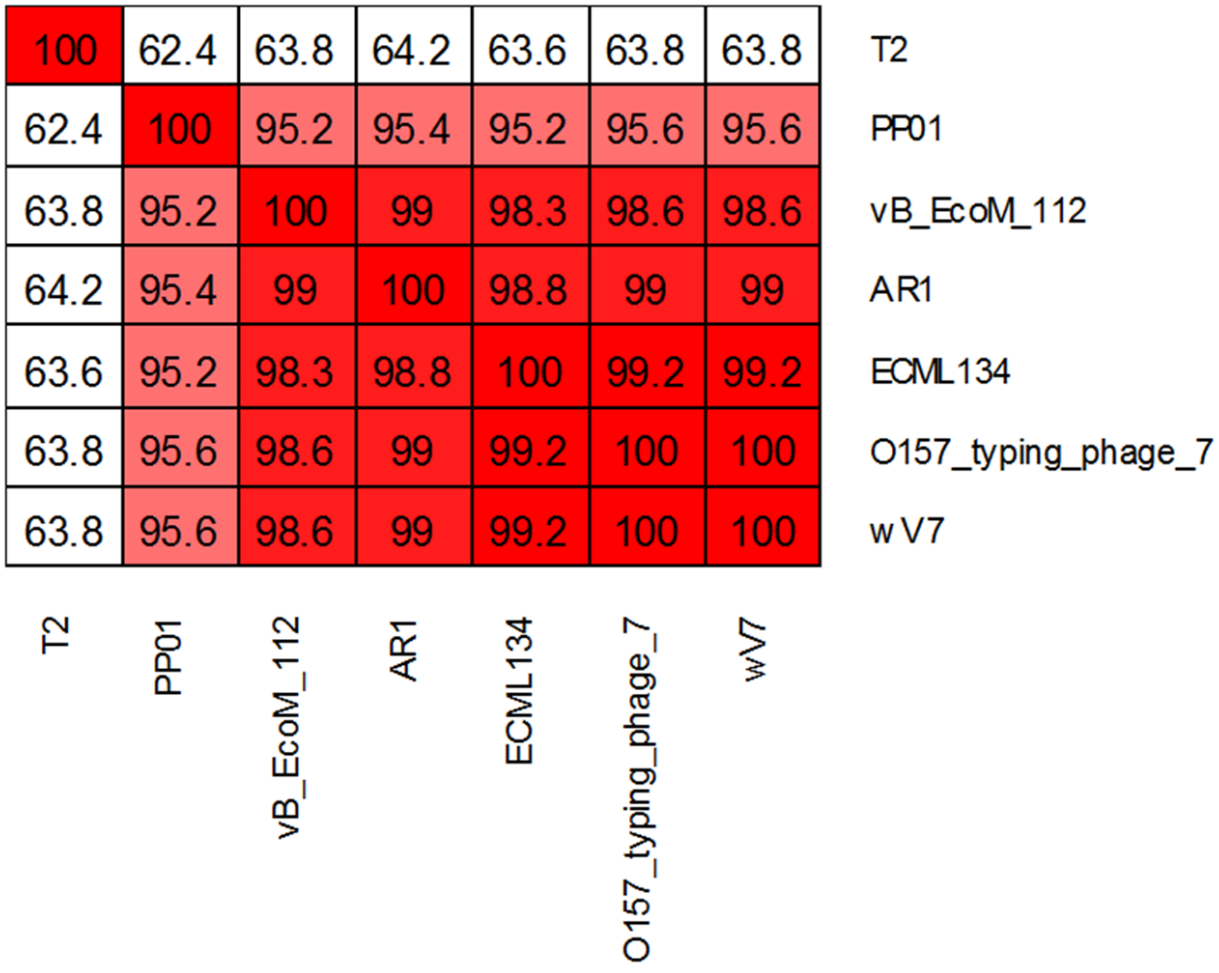


Fig. 2-8 Comparison of gp12 sequences among *E. coli* O157:H7 specific phage

The number in each cell is percentage of the amino acid identify between two indicated phages. The thicker red color represents the higher similarity.

CHAPTER 3 Screening of essential genes of PP01 for infecting *E. coli* O157:H7 by whole genome analysis

3.1 Introduction

As described above, the difference in infectivity of PP01 and T2 was not only due to their variation the in long tail fiber. In this chapter, comparative genome analysis of PP01 and T2 was conduct to identify PP01's essential genes for infection to *E. coli* O157:H7 besides the long tail fiber encoding genes. The genes highly variable between those phages were assumed to be candidates accounted for the difference of their infectivity. On top of that, quick screening using CRISPR/Cas9 was performed to further narrow down those candidates.

3.2 Material and method

3.2.1 Preparation of phage lysates and genomic DNA

Lysates of PP01 and T2 prepared by plate lysate method were further purified by CsCl density gradient centrifugation. Their genomic DNA was purified by phenol-chloroform extraction. The whole genomes were sequenced and assembled by 454 GS junior® and GS De Novo Assembler® (Roche).

3.2.2 NGS and data analysis

The genomes of PP01 and T2 were primarily annotated by transferring the annotation of T4 phage genome (NCBI accession no. AF158101) using the Genome Annotation Transfer Utility [43]. Gene prediction was also done by GeneMark.hmm with Heuristic models [44]. Those results were combined and curated manually. tRNA genes were

explored by tRNA-scanSE [45].

Pairwise alignment of PP01 and T2 genomes was conducted using Mauve [46]. Their sequence identity was calculated using 100 bp size of the sliding window at each base position by Perl script and plotted along with the alignment. The similarity plot and approximate positions of ORFs were drawn by Circos [47]. The homologous ORFs between PP01 and T2 were identified by reciprocal best hit Blast [48] using Blast program [49] and Perl script. The amino acid identities of the homologs were calculated by needle program in EMBOSS [38]. The genomic sequences of PP01 and T2 were deposited in the DDBJ database under accession number LC348379 and LC348380, respectively.

3.2.3 Essentiality test

28 of the ORFs highly variable between PP01 and T2 were picked up and pCas9-derived plasmids targeting those genes were constructed (Table 3-1). Then those plasmids were introduced into *E. coli* C. Their efficacy against PP01 phage was confirmed by measurement of EOP, using the same protocol described in section 2.3.4.

In order to test whether those 28 ORFs are essential for infection to *E. coli* O157:H7, 100 μl of the overnight culture of *E. coli* C carrying the pCas9-derived plasmids shown in Table 3-1 was inoculated to 2 ml LB medium (20 μg/ml Chloramphenicol) and incubated at 37°C until OD600 reached ~0.2. The phage lysate of PP01 was added to the host culture with MOI equal to ~0.01 and incubated for overnight to accumulate CEM (CRISPR Escape Mutant) in the population. Then the resultant lysate was recovered as supernatant by centrifugation (21600 g, room temperature, 5 min). This procedure was repeated and the obtained lysate was titrated using *E. coli* C and *E. coli* O157:H7. EOP value was calculated as the ratio of PFU number measured on *E. coli* O157:H7 to one

measured on *E. coli* C.

Table 3-1 Spacers for screening

Plasmid name	Target gene	Spacer sequence
pCas9-PP01_010	PP01_010	AATATTGTAAATCTCTCTAC
pCas9-motB	*motB*	AAATCCCGTTCTAAAGCTGC
pCas9-motB.1	*motB.1*	ATGACTACTGCTATTGAAAT
pCas9-motB.2	*motB.2*	GTGTTCAAATAACGCCATAC
pCas9-PP01_037	PP01_037	AACAAAATGTCTAAAGTATC
pCas9-PP01_038	PP01_038	GTTTAAATCAGAAGAAGATT
pCas9-PP01_048	PP01_048	AGCACCGTTATTCACACTAA
pCas9-ip10	*ip10*	TTTATTGCAGAATCTGCTGC
pCas9-PP01_068	PP01_068	TCTATTACTATTGAAAATTA
pCas9-PP01_076	PP01_076	AGCGGTATGTAGTAGAGAGT
pCas9-PP01_083	PP01_083	CTTTTCAGTATAAATTAATT
pCas9-49.2	*49.2*	AATTTTCTGAAAATGCATCA
pCas9-nrdC.4	*nrdC.4*	ACACTATTCTGACTAGACAT
pCas9-mobD.1	*mobD.1*	TTAGAAGTTAGTTCACTTTT
pCas9-tk.2	*tk.2*	CATGCTCATCTAGAAGTTCA
pCas9-tk.4	*tk.4*	ATGATTGTAAAATATATCAA
pCas9-vs.8	*vs.8*	GCATTCATCAGGGATAGATA
pCas9-e.7	*e.7*	TTAGATTCAAAAAGCTCTTC
pCas9-PP01_137	PP01_137	TCGGTCGGGGCTAAGCCTTG
pCas9-5.1	*5.1*	TGTTTCTAATTGAGCGGTAA
pCas9-hoc	*hoc*	ATAACCCCTGTTGGTGTTTT
pCas9-PP01_190	PP01_190	ATTTAGACTTTTAGCAGAAA
pCas9-30.4	*30.4*	AAATTAATTCAGAAATTATT
pCas9-30.8	*30.8*	AACGATTAATCTGAACGCTG
pCas9-PP01_223	PP01_223	CGAATGGTTTGAAGAAGATA
pCas9-cd.4	*cd.4*	CCTACTCTTCCATCAACAAT
pCas9-cd.5	*cd.5*	TGGCTTGGCATCATGAAACT
pCas9-ac	*ac*	ATGAATATTGCAAAATTATT

3.3 Results

3.3.1 General features of the genomes of PP01 and T2

The general genome features of PP01 and T2 were shown in Table 3-2. Their ORFs were annotated based on comparison with T4 genome. The genome of PP01 and T2 were highly similar to that of T4 as expected. Most of the functionally characterized genes of PP01 and T2 such as viral structural proteins and replication module were homologous with those of T4. The hypothetical genes which have no known function but are commonly conserved among T4-like phage [40] were also conserved in PP01 and T2.

Table 3-2 Summary of the whole genome data of PP01 and T2

Phage	Accession no.	Genome size	No. of ORFs	No. of tRNAs	GC%
PP01	LC348379	167812 bp	280	8	35.46%
T2	LC348380	163825 bp	276	9	35.35%

3.3.2 Overview of the genome comparison of PP01 and T2

The alignment of the genomic sequences of PP01 and T2 was drawn in Fig.3-1 with the plot of their similarity and the approximate positions of the ORFs to provide an overview of the genome comparison. Overall, their sequences were quite similar to each other as expected. Meanwhile, there were several highly variable regions such as the sequences around the genes encoding the long tail fiber (*36*, *37*, *38*) (90~95 kbp) and the whisker, baseplate, and short tail fiber (*wac*, *10*, *11*, *12*) (160~165 kbp). A few tens of insertions and deletions also existed. Their variations were especially abundant in the region where uncharacterized ORFs (highlighted by gray) were clustered (40~75 kbp).

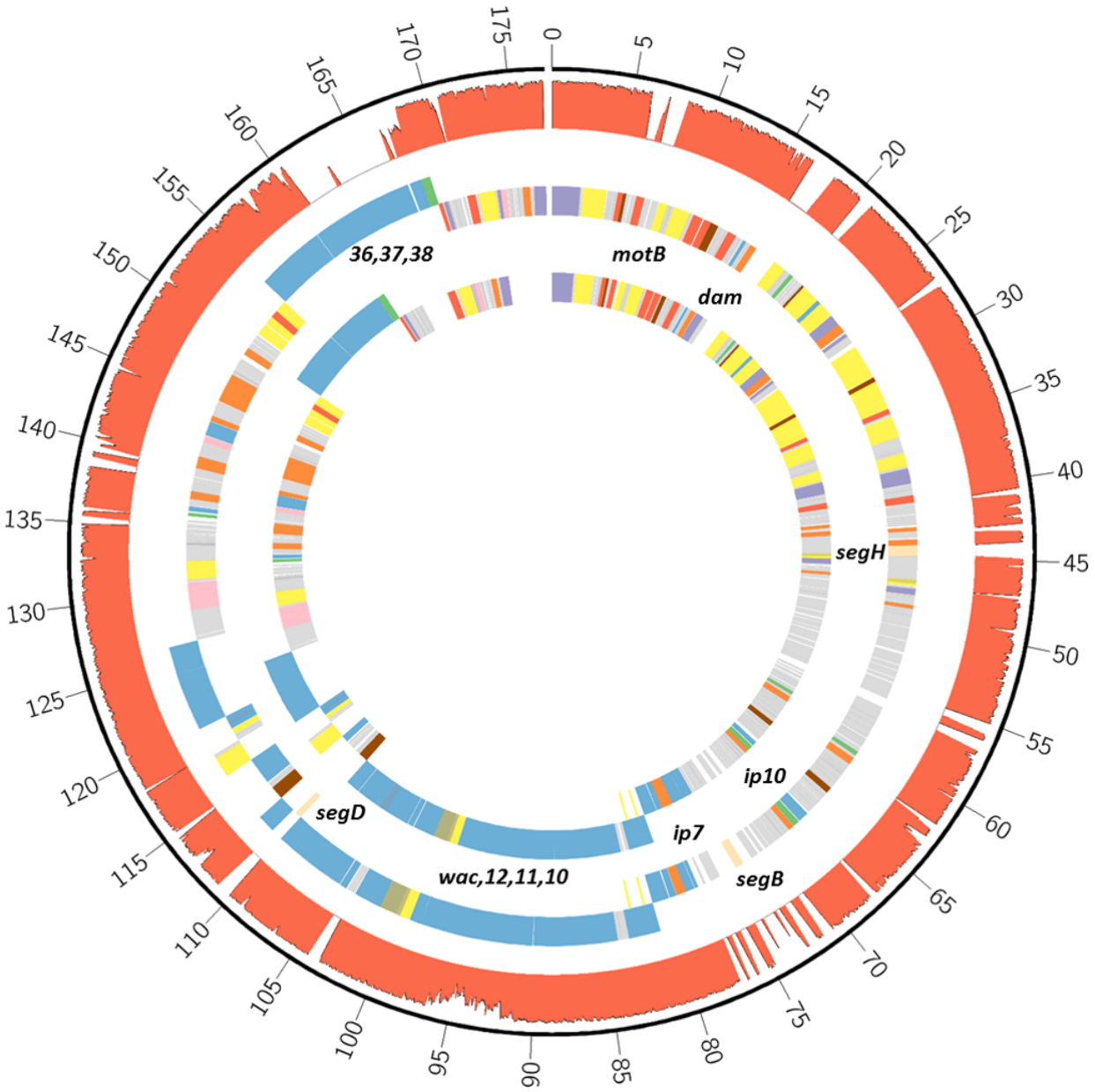

0
5
10
15
20
25
30
35
40
45
50
55
60
65
70
75
80
85
90
95
100
105
110
115
120
125
130
135
140
145
150
155
160
165
170
175
36,37,38
motB
dam
segH
ip10
ip7
segB
wac,12,11,10
segD

Fig 3-1 Overview of the whole genome comparison between PP01 and T2

The outermost circle (orange) represents the similarity plot (the height of the plot represents the height of the similarity). The approximate positions of ORFs of PP01 and T2 are shown by the outer and inner highlights, respectively. The same color scheme used in Miller et al. [20] was applied to show the function of the ORFs. Transcription: red; Translation: brown; Nucleotide metabolism: orange; DNA replication, recombination, repair, packaging, and processing: yellow; Virion proteins: blue; Lysis: green; Host or phage interactions: purple; Host alteration/shutoff: pink; Homing endonuclease: peach; Unknown: grey. The ticks along the circle show the length of sequence at each position (the unit is kbp).

3.3.2 Comparison of all ORFs of PP01 and T2

In order to compare the differences of their ORFs in detail, reciprocal best hit blast [48] was performed. The homologous ORFs which had a relatively lower similarity (less than 90%) are listed in Table 3-3, and the ORFs which were not shared between PP01 and T2 are also shown in Table 3-4. Summary of those results is displayed in Fig. 3-2.

207 of the ORFs were highly conserved (more than 90%) between PP01 and T2 and assumed to be unrelated to their difference of the infectivity. They included most of functionally characterized genes such as those encoding virion proteins, DNA replication modules and so on. Meanwhile, 52 of the homologs showed identity less than 90% and only 16 of those ORFs were functionally characterized (Fig. 3-2). Those 17 ORFs were annotated as the genes related to virion (*soc*, *hoc*, *10*, *11*, *12*, *wac*, *36*, *37*, *38*), transcription (*motB*), translation (*modB*, *rnlB*), nucleotide metabolism (*pseT*), DNA replication (*repEA*), lysis (*sp*) and host-phage interactions (*ac*, *pin*) [20].

Besides, PP01 and T2 had 21 and 17 ORFs not shared with each other, respectively. Only a few of them were known about their function (Fig.3-2). The gene *dam* encodes DNA-adenine methylase, which methylates N6 of adenine in genomic DNA. T2 had this gene as already reported [50], whereas PP01 lacked it. Thus PP01 seemed to have genomic DNA without methylation of adenine, unlike T2. The gene *nudE* is Nudix hydrolase which active on FAD, Ap3A, and ADP-ribose [51]. Although the regions encoding *nudE* of PP01 and T2 were quite similar, the T2 *nudE* had one base deletion resulting in frame-shift. I reasoned this was probably a spontaneous mutation introduced through the propagation in our lab, based on comparison with another report of the sequence of T2 *nudE* (GenBank accession number AYD82726.1). PP01 had genes *segB*, *segD*, and *segH*, which are mobile genetic elements commonly appeared in T4-like phages [52], [53]. In addition, one hypothetical mob-like homing endonuclease could be found in PP01 and T2 genomes (locus-tag: PP01_057 and T2_056). IP (internal protein) is a small peptide which is incorporated into the head protein during assembly of the virus particle and injected into the host cell along with genome DNA upon infection [54], [55]. IP4, IP5, and IP7 were present in T2 genome. PP01 shared IP4 and IP5 with T2 but had IP10 in place of IP7. Since the incorporation of IP into the capsid is strictly dependent on a conserved capsid targeting sequence located in the N-terminal portion [54], [55], searching this consensus sequence confirmed they were only IP existing in PP01 and T2. IP is consider to play important roles in the early stage of infection to counteract the host restriction system or subvert host metabolism. In fact, IP1 can inactivate specific host restriction enzymes to protect co-injected genomic DNA [56]. However, the functions of the other IPs remain to be characterized.

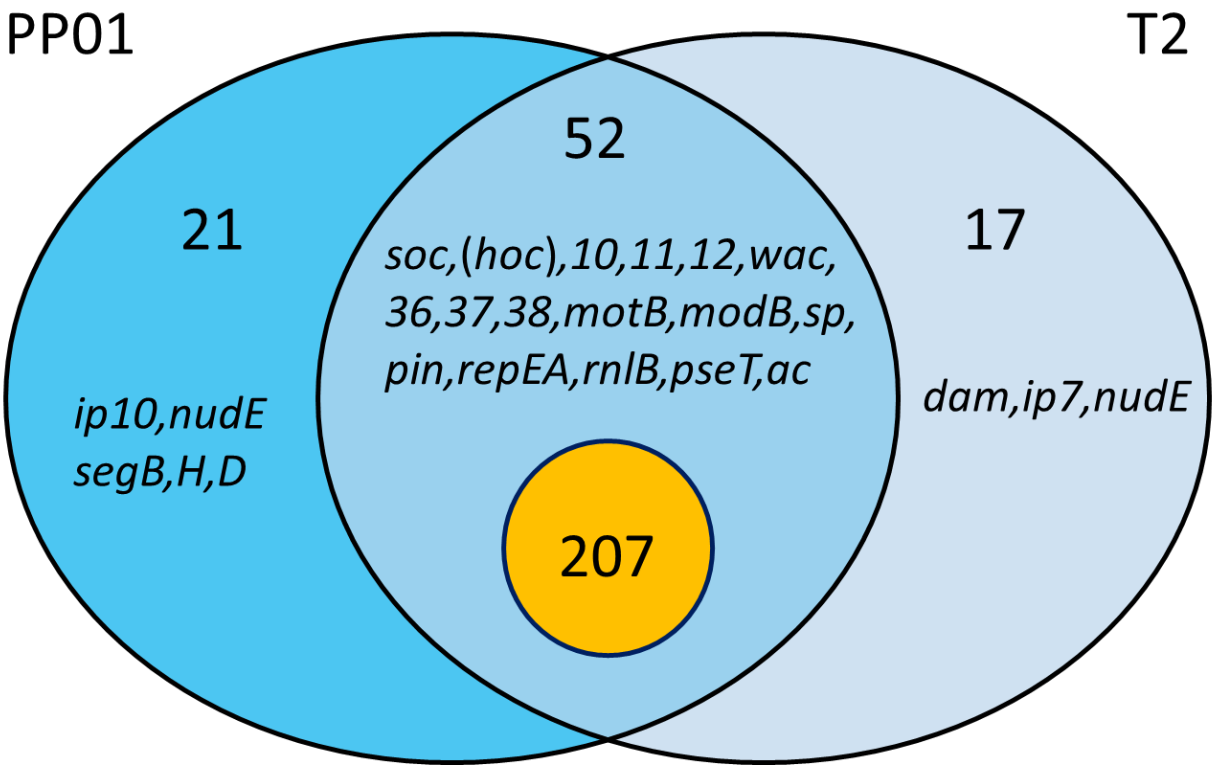


Fig. 3-2 Summary of the comparison of ORFs between PP01 and T2

Summary of the comparison of ORFs between PP01 and T2 was shown by Venn diagram. The subset of the homologs which had more than 90 % identity are represented by the yellow circle in the intersection. In the other region, the names of the functionally characterized genes are shown by the letters.

Table 3-3 Homologs between PP01 and T2 which showed less than 90% of amino acid identity

gene	PP01 locus tag	T2 locus tag	Identity of amino acid sequence (%)
motB	PP01_011	T2_010	47
dexA.2	PP01_016	T2_013	89
modB	PP01_021	T2_019	62.5
soc	PP01_029	T2_027	78.2
—	PP01_031	T2_029	71.6
61.2	PP01_035	T2_035	86.6
sp	PP01_036	T2_036	88.8
61.4	PP01_039	T2_037	87.2
—	PP01_057	T2_056	69.8
55.3	PP01_071	T2_069	68.7
55.6	PP01_074	T2_072	63.9
pin	PP01_084	T2_079	85.8
49.2	PP01_086	T2_081	58.9
nrdC.2	PP01_091	T2_086	57
nrdC.4	PP01_093	T2_088	44.6
nrdC.5	PP01_094	T2_090	77.8
nrdC.6	PP01_095	T2_092	62.9
nrdC.7	PP01_096	T2_093	61.2
—	PP01_101	T2_099	77.4
—	PP01_102	T2_101	28.2
mobD.1	PP01_103	T2_102	37.6
mobD.2	PP01_104	T2_103	88.9
tk.2	PP01_113	T2_113	56.5
tk.4	PP01_117	T2_117	56.2
vs.8	PP01_126	T2_126	35.4
trna.2	PP01_139	T2_139	84.3
repEA	PP01_151	T2_152	71.2
5.1	PP01_152	T2_153	48.5
10	PP01_158	T2_159	85.1
11	PP01_159	T2_160	68.6

12	PP01_160	T2_161	62.4
wac	PP01_161	T2_162	80.5
rnlB	PP01_182	T2_181	70.9
hoc	PP01_185	T2_184	29
30.3'	PP01_210	T2_209	68.4
30.4	PP01_211	T2_210	33.3
cd.4	PP01_225	T2_222	24.6
pseT	PP01_227	T2_223	77
—	PP01_228	T2_224	75.7
pseT.3	PP01_231	T2_227	89.8
frd.2	PP01_243	T2_239	81.1
36	PP01_252	T2_248	80.9
37	PP01_253	T2_249	28.7
38	PP01_254	T2_250	48.2
asiA.1	PP01_257	T2_253	86.3
arn.3	PP01_262	T2_258	75.6
motA.1	PP01_265	T2_261	88
ac	PP01_268	T2_264	26.4
stp	PP01_269	T2_265	75
ndd.2a	PP01_273	T2_269	61
—	PP01_277	T2_273	67.6

Table 3-4 ORFs uncommon between PP01 and T2

(A) The ORFs which existed only in PP01

locus tag	gene name	Best hit against nr
PP01_010	—	hypothetical protein RB51ORF010 [Enterobacteria phage RB51]
PP01_012	*motB.1*	Hypothetical protein vBEcoMUFV13_g011 [Escherichia phage vB_EcoM-UFV13]
PP01_013	*motB.2*	putative MotB.2 [Shigella phage Shfl2]
PP01_037	—	hypothetical protein vBEcoMUFV13_g039 [Escherichia phage vB_EcoM-UFV13]
PP01_038	—	hypothetical protein [Clostridioides difficile]
PP01_048	—	hypothetical phage protein [Escherichia phage wV7]
PP01_068	—	hypothetical protein [Clostridioides difficile]
PP01_076	—	hypothetical protein PhAPEC2_71 [Escherichia phage vB_EcoM_PhAPEC2]
PP01_079	*segH*	SegH [Enterobacteria phage RB3]
PP01_083	—	hypothetical protein HY01_0073 [Escherichia phage HY01]
PP01_128	*ip10*	hypothetical protein [Clostridioides difficile]
PP01_130	*nudE*	nudix hydrolase [Yersinia phage PST]
PP01_135	*e.7*	e.7 hypothetical protein [Enterobacteria phage T4]
PP01_137	—	unnamed protein product [Escherichia phage slur02]
PP01_138	*segB*	hypothetical protein [Salmonella enterica]
PP01_171	—	hypothetical protein ECML134_163 [Escherichia phage ECML-134]
PP01_180	*segD*	endonuclease [Enterobacteria phage T4]
PP01_190	—	hypothetical protein e112_194 [Escherichia phage vB_EcoM_112]
PP01_214	*30.8*	gp30.8 hypothetical protein [Enterobacteria phage RB51]
PP01_223	—	hypothetical protein RB32ORF214c [Enterobacteria phage RB32]
PP01_226	*cd.5*	hypothetical protein RB51ORF220 [Enterobacteria phage RB51]

(B) The ORFs which existed only in T2

locus tag	gene name	Best hit against nr
T2_018	—	hypothetical protein ECML134_019 [Escherichia phage ECML-134]
T2_031	*dam*	DNA adenine methyltransferase [Escherichia coli]
T2_032	—	hypothetical protein e112_034 [Escherichia phage vB_EcoM_112]
T2_046	—	hypothetical protein RB27_048 [Enterobacteria phage RB27]
T2_055	—	homing endonuclease [Shigella phage SHFML-11]
T2_089	*nrdC.4*	thioredoxin, phage-associated [Yersinia phage phiD1]
T2_091	—	hypothetical protein nrdc.6 [Shigella phage Shfl2]
T2_097	—	hypothetical protein [Salmonella enterica]
T2_100	—	hypothetical protein [Escherichia phage ECO4]
T2_112	*tk.1*	hypothetical protein RB32ORF108c [Enterobacteria phage RB32]
T2_128	*nudE*	MULTISPECIES: soluble lysozyme [Bacteria]
T2_134	—	hypothetical protein RB14ORF134 [Enterobacteria phage RB14]
T2_136	—	trna.2 protein [Enterobacteria phage T2L]
T2_137	—	trna.1 protein [Enterobacteria phage T2L]
T2_138	—	hypothetical protein RB27_137 [Enterobacteria phage RB27]
T2_141	—	Ip7 protein - phage T2
T2_185	—	terminase large subunit [Escherichia phage ECML-134]

3.3.3 Screening of essential gene of PP01 for infecting *E. coli* O157:H7 by CRISPR/Cas9

The comparison of the ORFs between PP01 and T2 showed that among PP01's ORFs, 52 were relatively distinct from T2's ORFs and 21 were not shared with T2 (Fig. 3-2). In order to further narrow down those 73 candidates, 28 ORFs were arbitrarily chosen and their essentiality for infection to *E. coli* O157:H7 was assessed using CRISPR/Cas9. 12 ORFs whose identities to T2's homologs were less than 50% were chosen from the 52 ORFs common between PP01 and T2, and 16 ORFs were selected from the 21 ORFs unique in PP01 (Table 3-1).

When phage is grown on host bacteria carrying CRISPR/Cas, after several generation a mutant phage which spontaneously lost protospacer sequence (CEM; CRISPR Escape Mutant) can be dominant in the population due to growth-inhibition by CRISPR/Cas [57]. That mutation can cause loss of function of the ORF coding the protospacer by frame-shift or alteration of amino acid residue. Thus lysate of PP01 propagated using *E. coli* C carrying CRISPR/Cas9 should dominantly contain the mutant which lost the function of the targeted gene. If that gene is essential for infection of *E. coli* O157:H7, the lysate should show significantly reduced titer on *E. coli* O157:H7 compared to titer on *E. coli* C.

Firstly, EOP was measured to confirm pCas9-derived plasmids (Table 3-1) could inhibit the infection of PP01 phage (Fig. 3-3). In all cases, infection of PP01 was inhibited. The EOP values were mostly ~10^{-1} and comparable with the case of T2 phage, implying CRISPR/Cas9 could also work for gene manipulation of PP01.

Thus the screening was performed using those plasmids (Fig. 3-4). The lysate of PP01 propagated using *E. coli* C/pCas9-PP01-motB showed a significant decrease of titer (~40 fold) on *E. coli* O157:H7. For the other cases, the titers were reduced ~2 to ~7 fold but the reduction was comparable with one of negative control (pCas9). The mutant of PP01 isolated from the co-culture with *E. coli* C/pCas9-PP01-motB showed ~10^{-6} fold less titer on *E. coli* O157:H7 than one on *E. coli* C. Besides, partial deletion of its *motB* gene was confirmed by Sanger sequencing (data not shown). These results implied that PP01 *motB* gene was essential for infection of *E. coli* O157:H7.

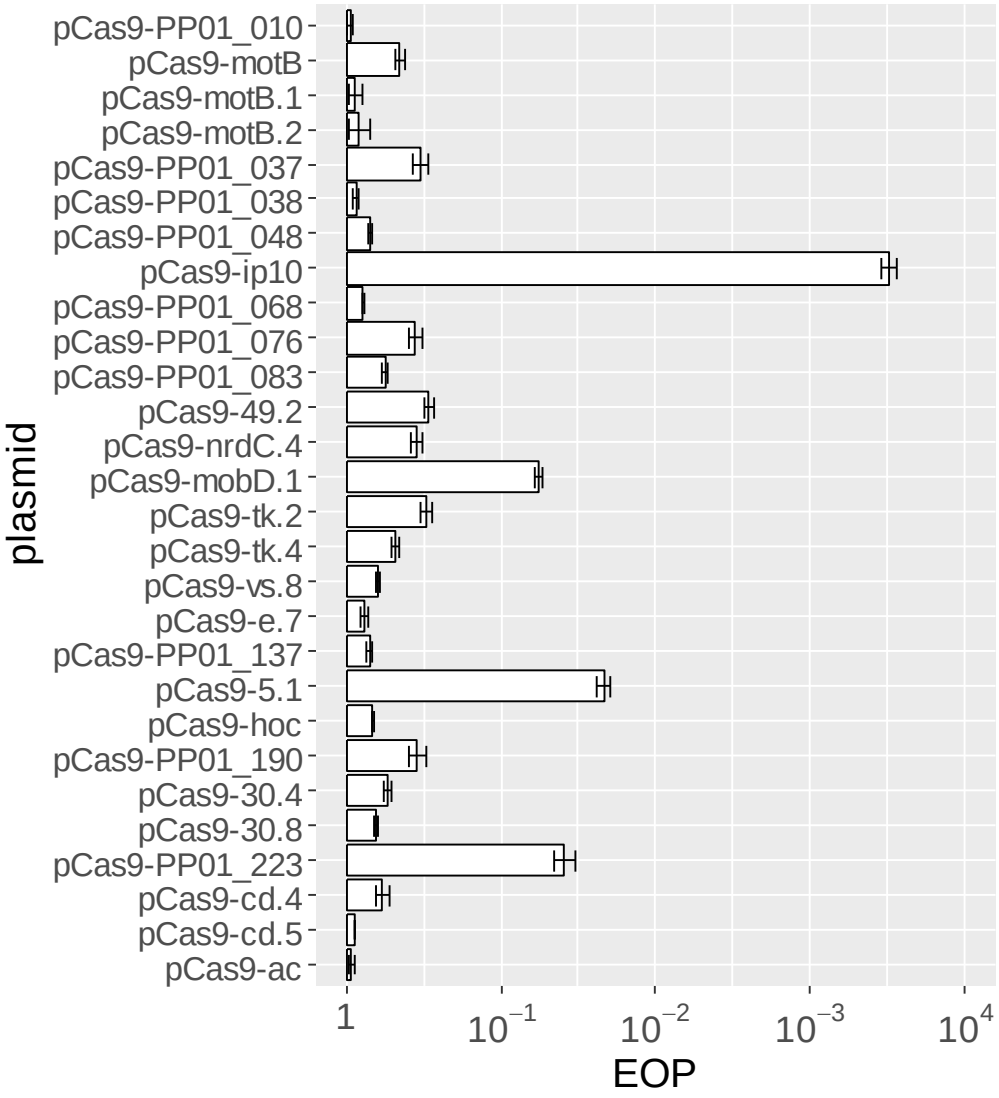


Fig. 3-3 Efficacy of CRISPR/Cas9 against PP01 phage

EOP was defined as the ratio of phage titer estimated on the host (*E. coli* C) carrying CRISPR/Cas9 to the host without CRISPR/Cas9. Thus the reduction of EOP represents the extent of inhibition of infection by CRISPR/Cas9. The data represent means ± standard deviations of three technical replicates.

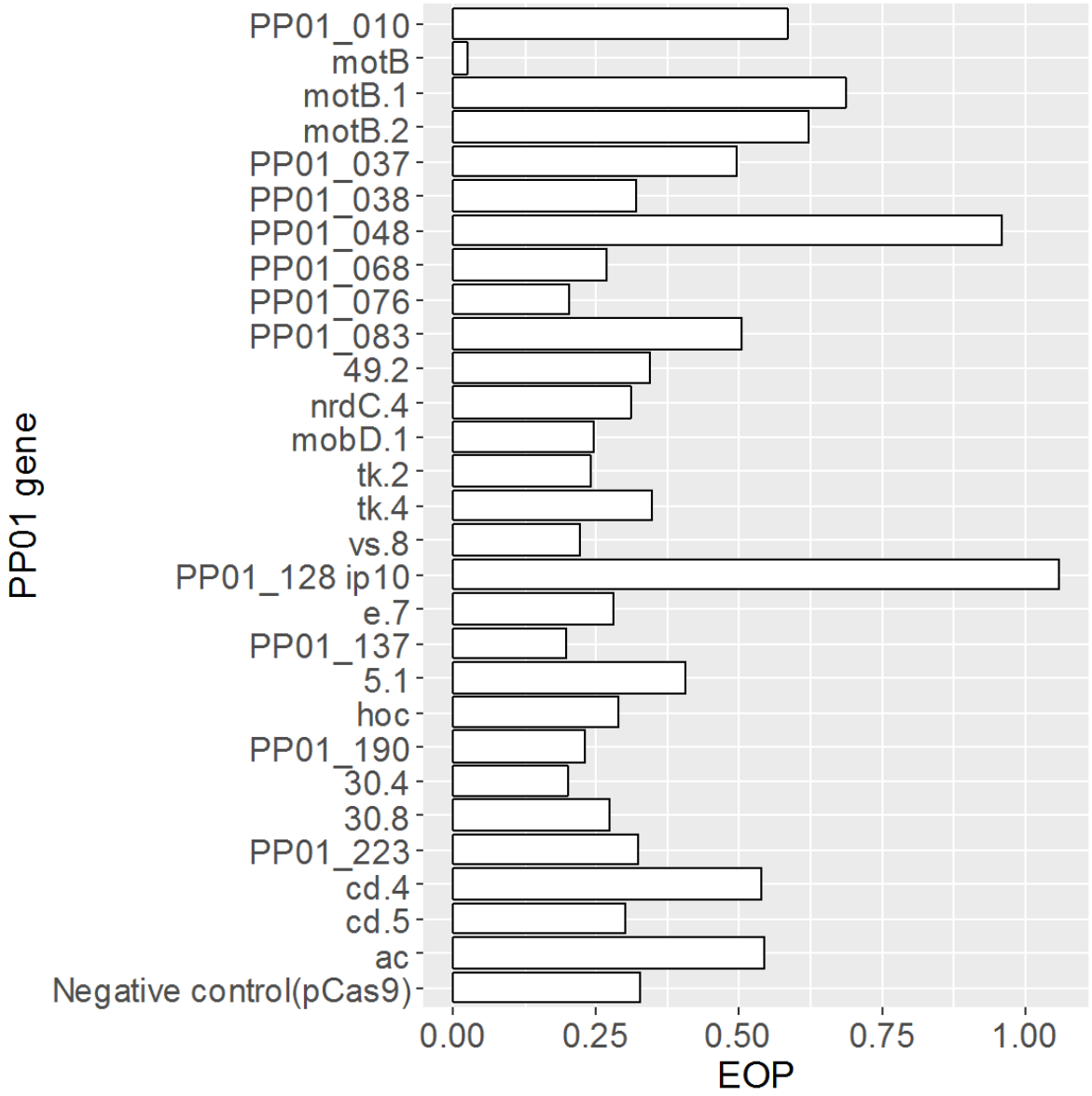


Fig. 3-4 Result of screening of essential gene for infection to *E. coli* O157:H7

PP01 was co-cultured with *E. coli* C carrying CRISPR/Cas9 targeting each of the ORFs indicated in the Y-axis, possibly resulting in loss of that ORFs. X-axis shows EOP values (the titer on *E. coli* O157 divided by one on *E. coli* C) of each of the resultant lysates. The reduction of EOP implies loss of the targeted gene could severely reduce the infectivity of PP01 to *E. coli* O157:H7.

3.4 Discussion

3.4.1 The dissimilarity between PP01 and T2 in the long tail fiber, the short tail fiber and the outer capsid proteins

As shown above, the dissimilarity of gene *37* and *38*, which encodes distal long tail fiber, was responsible for the difference of adsorption ability between PP01 and T2. gp37 and gp38 are connected to the virion through gp34, gp35, and gp36 (Fig. 1-3). Whereas the amino acid sequences of gp34 and gp35 were almost identical (around 95% of identity) between PP01 and T2, gp36 (80.9% of identify) were relatively variable. gp36 perhaps needed to be modified to connect highly diverged gp37 to gp35. The gene *wac* encodes the whisker of T4-like phage, which plays a role in regulating retraction of the long tail fiber and efficient attachment of the long tail fiber to the phage particle during phage assembly [58]. Hence gene *wac* might be also mutated to adopt diversification of the long tail fiber.

Similarly, most of the components of the baseplate (gp7, 8, 9, 10, 11, 12, 6, 25, 53, 5, 27, 29, 26, 28, 48 and 54) [40] of PP01 and T2 were highly conserved, whereas gp10, gp11, and gp12 were quite varied (85.1%, 68.6 and 62.4% of identity, respectively). The variation of gp12 was shown to be responsible for the difference of their adsorption ability to the host in chapter 2. Among T4-like phage, gp12 is connected to the baseplate through gp10 and gp11. Therefore those of PP01 might be modified to compensate diversification of gp12 and enable to interact with it. This could be the reason why homologous recombination of entire gp12 between PP01 and T2 didn't work (data not shown). Exchange of N-terminal part of their gp12 could disturb the interaction between gp12 and the base plate, resulting in failure of forming a complete phage particle. Since the variations of gene *10*, *11*, *wac*, *36* between PP01 and T2 were corresponding to their

dissimilarity of the long and short tail fibers, they were reasonably presumed not to be accounted for the difference of infectivity between PP01 and $T2_{PP01\text{-}g37\text{-}38}$.

Meanwhile, PP01's *soc* and *hoc* genes distinct from T2 might be worthy to analyze in further research. Hoc (highly antigenic outer capsid protein) is virion protein, which decorates the outer-surface of the capsid of T4-like bacteriophage [59]. In T4 phage, Hoc is composed of 4 domains, of which domain 1, 2 and 3 have an immunoglobulin-like fold. Other T4-like phages typically have 1~5 domains [60], although T2 phage is known not to express Hoc [61]. The sequences of T4 *hoc* gene and the locus of T2 genome corresponding to it were highly similar to each other, but the T2 *hoc* ORF was truncated by several insertions and deletions. Glu-Ser-Arg-Asn-Gly in the C-terminal domain of Hoc is strictly conserved in all-known T4-like phage and considered to be necessary for the binding of Hoc to the capsid [60]. Since this part was missing in the truncated Hoc ORF of T2, even though this ORF could be expressed, it shouldn't properly attach to the T2 capsid. It was consistent with the fact that T2 does not have Hoc. Meanwhile, Hoc of PP01 have 5 domains and domain 2 and 3 were almost identical and seemed to be duplicated. Those domains were quite similar to the counterpart of Hoc in T4. Since, Hoc is connected to the head protein through Soc, the dissimilarity of Soc between PP01 and T2 might be due to the differences in Hoc proteins. It is suggested Hoc can contribute to improved fitness by attaching phage capsid to the surface of *E. coli* via Hoc [60]. Thus it should be interesting to study the role of PP01's Hoc in infectivity against *E. coli* O157:H7 based on comparison with T2 phage, which had no Hoc.

3.4.2 Uncertainty of the essentiality of *motB* for infection to *E. coli* O157:H7

The essentialities of PP01's genes were quickly surveyed exploiting occurrence of CEM under the pressure of CRISPR/Cas9. Although the result implied *motB* was important for PP01 to infect *E. coli* O157:H7, it is still uncertain whether the genes other than *motB* are not essential for the infectivity. Firstly, since the mutation to escape CRISPR/Cas9 can be large deletion ranging more than one ORFs, so the neighbor of *motB* such as *cef* and *motB.2* might be also destroyed during the co-culture of PP01 and *E. coli* C/pCas9-motB. Secondly, the result of screening possibly includes false positives. Although the EOP test showed that CRISPR/Cas9 could inhibit the infection of PP01 phage, its efficiency was varied depending on the spacer sequences and not high for most of the targets, similarly to the case of T2 phage. Therefore it is possible CRISPR/Cas9 didn't work well for some genes enough to accumulate of CEM and evaluate the essentiality of the target gene by EOP assay. In order to make screening process quick, it was not confirmed if CEM actually dominated in each co-culture. These problems will be addressed in the next chapter.

CHAPTER 4 Modification of *motB* of PP01 and T2

4.1 Introduction

motB gene of PP01 was implied to be essential for its distinctive infectivity to *E. coli* O157:H7, but it still remained unclear. In this chapter, to elucidate difference of infectivity between PP01 and $T2_{PP01\text{-}g37\text{-}38}$ is accounted for the variation in *motB*, firstly, clean knock-out of PP01's *motB* was tried to construct using CRISPR/Cas9 system, which was shown to work for T2 phage. Secondly *motB* of $T2_{PP01g37\text{-}38}$ phage was replaced with one of PP01 to gauge whether the resulting recombinant can acquire infectivity to *E. coli* O157:H7. In addition, phylogenetic analysis of *motB* of T4-like phages was conducted to get more insight about this gene.

4.2 Material and methods

4.2.1 Bacterial and phage strains

Two host bacteria (*E. coli* C and *E. coli* O157:H7 (ATCC43888)) and two wild type phages (PP01 and T2) were exploited also for the study in this chapter (See section 2.2.1 for details). In addition, Three mutant phages were used: $T2_{PP01g37\text{-}38}$ (described in the chapter 2), $T2_{PP01g37\text{-}38\text{-}motB}$ and, PP01ΔmotB (constructed as follows). The used host and phage strains were summarized in Table 4-1.

Table 4-1 Bacteria and phage strains

Strain name	Description
Bacteria	
E. coli C	Susceptible to both PP01 and T2
E. coli O157:H7 (ATCC43888)	Susceptible to PP01
Bacteriophage	
$T2_{PP01g37\text{-}38}$	T2 recombinant with PP01's long tail fiber (gene *37* and *38*)
$T2_{PP01g37\text{-}38\text{-}motB}$	T2 recombinant with PP01's long tail fiber and *motB* gene
PP01	O157:H7 phage
PP01ΔmotB	PP01 *motB* knock-out mutant

4.2.2 Editing templates

The editing templates to delete PP01's *motB* and replace *motB* of $T2_{PP01g37\text{-}38}$ with the counterpart of PP01 were constructed as follows using primers shown in Table 4-2. For deletion, the editing template was constructed by overlap PCR. The first PCRs were conducted using PP01 genome as template and primer pairs motB-del-up-f/motB-del-up-r and motB-del-dw-f/motB-del-dw-r to amplify the flanking regions of PP01 motB. In the second PCR, those amplicons were used as templates of PCR with primer pair motB-del-up-f/mot-del-dw-r and combined due to 15 bp overlap region added to 5' end of motB-del-up-r and motB-del-dw-f (Fig 4-1). For replacement, since 53 bp of N-terminal parts of PP01 and T2 *motB* were identical and the rest parts were highly variable, the primer pair motB-ex-f/motB-ex-r was designed to amplify the entire ORF of PP01 *motB*. Besides, 50 bp nucleotides homologous with the 3' flanking region of T2 *motB* was added to 5' end of the reverse primer. Thus the amplicon of PCR with these primers resulted in the variable region of PP01 which was adjoined by 5' and 3' flanking regions of the variable region of T2 (Fig. 4-1). The obtained DNA fragments were digested with EcoRI and BamHI (New England Biolab) and ligated with pUC118 which was digested with the same enzymes. The plasmids for deletion and replacement were named pD-motB and pEx-motB, respectively.

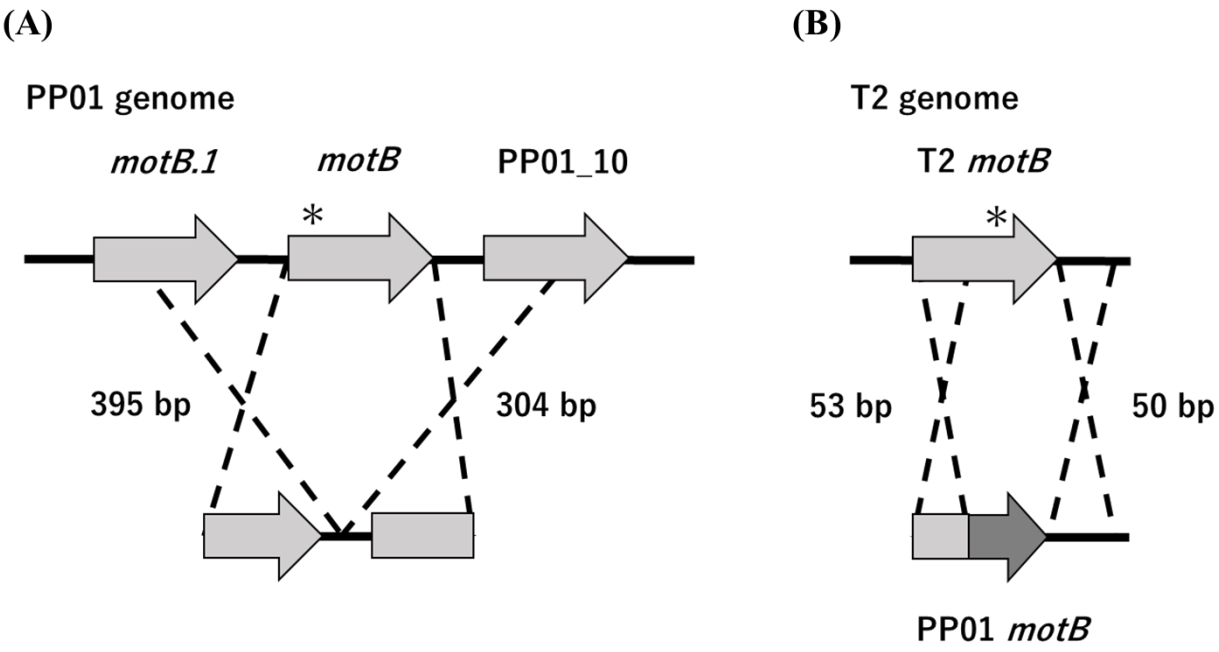


Fig. 4-1 Structure of editing templates

The structure of the editing templates to modify *motB* of PP01 (A), and *motB* of T2 (B) are displayed. The arrows indicate the ORFs of PP01 and T2. The dashed lines represent the regions homologous between the editing templates and the phage genome. The numbers indicate the length of those regions. The asterisks show the approximate positions of the proto-spacer sequences.

Table 4-2 Primers for construction of the editing templates

Primer name	Target	Sequence
motB-ex-f	pEx-motB	AGAGGATCCATGATTATTAATATTGGTGAAATTGCTC
motB-ex-r	pEx-motB	TACGAATTCACCATAATCAACTACTGATGTATATAGTTTTAT GAAAAAATTTTTAAACTTTAAGCTTTACGAGGAGAAGC
motB-del-up-f	pD-motB	CATGATTACGAATTCCGACTTGATTCGTTGTCCATG
motB-del-up-r	pD-motB	GGAATTATTATGAATATTGTAAATCTCTCTACTGG
motB-del-dw-f	pD-motB	ATTCATAATAATTCCAGTTATCTCTCACTTGTTAAAAAGAT TTTATACTC
motB-del-dw-r	pD-motB	GACTCTAGAGGATCCAGGCGAATTTACCGTTGAAG

4.2.5 Homologous recombination and screening of recombinant phage

The homologous recombination experiment to modify *motB* gene of PP01 and T2 was conducted with the basically same protocol described in 2.2.5. The hosts and the plasmids used for recombination and screening were shown in Table 4-3. In brief, *E. coli* C was transformed by the plasmids carrying editing template and CRISPR/Cas9 (for deletion, pD-motB and pCas9-PP01-motB; for replacement, pEx-motB and pCas9-T2-motB). Those strains were used as host to grow PP01 or $T2_{PP01\text{-}g3738}$ for two passages and the obtained lysates were plated on the hosts for screening. The *motB* replaced recombinant of $T2_{PP01g37\text{-}38}$ was screening by plating on *E. coli* O157:H7 since initial attempt to select by CRISPR/Cas9 failed by uncertain reason. The *motB* sequence of the resultant plaques were checked by plaque direct PCR and Sanger sequencing using primer pairs shown in Table 4-4.

Table 4-3 Plasmids and host bacteria for the recombination experiments

Plasmid name		
CRISPR/Cas9	Target	Spacer sequence
pCas9-PP01-motB	PP01 *motB*	AAATCCCGTTCTAAAGCTGC
pCas9-T2-motB	T2 *motB*	GGTTTAGACATTCCATCTTT
Editing template		
pD-motB	For knock-out of PP01 *motB*	
pEx-motB	For replacement of T2 *motB* with PP01 *motB*	

Combinations of the hosts and the plasmids for the recombination experiment			
Modified gene	Host	Recombination	Selection
motB of PP01	C	pCas9-PP01-motB & pD-motB	pCas9-PP01-motB
motB of $T2_{PP01\text{-}g3738\text{-}motB}$	C	pCas9-T2-motB & pEx-motB	pCas9-T2-motB (and *E. coli* O157:H7)

Table 4-4 Sequences of the primers for screening of recombinants

Primer name	Target	Sequence
PP01-motB-seq-f	PP01 motB	CTTTAGAACGAGAAGCAACGGATAC
PP01-motB-seq-r	PP01 motB	GTCAGAGGAAGAGTGCGATC
T2-motB-seq-f	T2 motB	GTTTCAGAAGAGTAGCAGCACC
T2-motB-seq-r	T2 motB	TTTACTTCCTCGGTTAGTTGTGG

4.2.6 Infectivity test

The infectivity of *motB* mutant of PP01 and $T2_{PP01\text{-}g37\text{-}38}$ to *E. coli* O157:H7 was evaluated by EOP (the ratio of the titer measured on *E. coli* O157:H7 to one evaluated on *E. coli* C). They were grown using *E. coli* C or *E. coli* O157:H7 to see whether host for propagation affected the infectivity of the progeny.

4.2.7 Phylogenetic analysis of *motB* of T4-like virus

The genome sequences of T4-like phages were downloaded from NCBI taxonomy database [62]. The data duplicated or unannotated were excluded. The accession numbers of those sequences were listed in Table 4-5. *motB* of 66 T4-like phages were extracted

from their genomic data. Multiple alignment was performed using ClustalW [63] and phylogenetic tree was drawn by Biophyhon Phylo module [64].

Table 4-5 T4-like phages for phylogenetic analysis of *motB*

Phage	Accession number
Bacteriophage_RB32	DQ904452
Citrobacter_phage_vB_CroM_CrRp10	MG775043
Enterobacteria_phage_AR1	AP011113
Enterobacteria_phage_RB10	KM606999
Enterobacteria_phage_RB14	FJ839692
Enterobacteria_phage_RB18	MH553563
Enterobacteria_phage_RB27	KM607000
Enterobacteria_phage_RB3	KM606994
Enterobacteria_phage_RB33	KM607001
Enterobacteria_phage_RB5	KM606995
Enterobacteria_phage_RB51	FJ839693
Enterobacteria_phage_RB55	KM607002
Enterobacteria_phage_RB59	KM607003
Enterobacteria_phage_RB6	KM606996
Enterobacteria_phage_RB68	KM607004
Enterobacteria_phage_RB7	KM606997
Enterobacteria_phage_RB9	KM606998
Enterobacteria_phage_T4	AF158101
Enterobacteria_phage_T6	MH550421
Enterobacteria_phage_ime09	JN202312
Enterobacteria_phage_vB_EcoM_ACG-C40	JN986846
Enterobacteria_phage_vB_EcoM_IME339	MH051915
Enterobacteria_phage_vB_EcoM_IME340	MH051916
Escherichia_coli_O157_typing_phage_7	KP869105
Escherichia_phage_EC121	MF001359
Escherichia_phage_ECML-134	JX128259
Escherichia_phage_ECO4	MF001360
Escherichia_phage_HY01	KF925357
Escherichia_phage_HY03	KR269718
Escherichia_phage_PE37	KU925172
Escherichia_phage_PEC04	KR233165
Escherichia_phage_PP01	LC348379
Escherichia_phage_T2	LC348380
Escherichia_phage_UFV-AREG1	KX009778
Escherichia_phage_YUEEL01	KY290975
Escherichia_phage_e11_2	KJ668714
Escherichia_phage_slur02	LN881726

Escherichia_phage_slur03	LN881728
Escherichia_phage_slur04	LN881729
Escherichia_phage_slur07	LN881732
Escherichia_phage_slur08	LN881733
Escherichia_phage_slur11	LN881734
Escherichia_phage_slur13	LN881737
Escherichia_phage_slur14	LN881736
Escherichia_phage_vB_EcoM-UFV13	KU867876
Escherichia_phage_vB_EcoM-fFiEco06	MG781190
Escherichia_phage_vB_EcoM-fHoEco02	MG781191
Escherichia_phage_vB_EcoM_JB75	MH355584
Escherichia_phage_vB_EcoM_NBG2	MH243439
Escherichia_phage_vB_vPM_PD112	MH837626
Escherichia_phage_wV7	HM997020
Salmonella_phage_SG1	MF001354
Shigella_phage_SH7	KX828711
Shigella_phage_SHBML-50-1	KX130864
Shigella_phage_SHFML-11	KX130861
Shigella_phage_SHFML-26	KX130862
Shigella_phage_Sf21	MF327007
Shigella_phage_Sf22	MF158045
Shigella_phage_Sf23	MF158046
Shigella_phage_Sf24	MF327008
Shigella_phage_Sf25	MF327009
Shigella_phage_Shfl2	HM035025
Shigella_phage_pSs-1	KM501444
Staphylococcus_phage_SAJK-IND	MG010123
Yersinia_phage_PST	KF208315
Yersinia_phage_phiD1_complete_genome	HE956711

4.3 Result

4.3.1 Construction and characterization of PP01ΔmotB and T2$_{PP01\text{-}g3738\text{-}motB}$

The application of CRISPR/Cas9 to genome editing of PP01 was successful as in the case of T2 phage. The *motB* knock-out mutant of PP01 was obtained upon only screening of 10 plaques and named PP01ΔmotB. Meanwhile, the first attempt to screen the T2 *motB* recombinant by CRISPR/Cas9 was failed, despite the fact that pCas9-T2-motB was decently working (Fig. 4-2). This might be because T2 *motB* was beneficial to T2 phage to infect *E. coli* C and loss of *motB* was more deleterious than the pressure of

CRISPR/Cas9. Thereby screening was performed by plating on *E. coli* O157:H7. The isolated recombinant was named $T2_{PP01\text{-}g37\text{-}38\text{-}motB}$.

The infectivity of PP01ΔmotB and $T2_{PP01\text{-}g37\text{-}38\text{-}motB}$ to *E. coli* O157:H7 was assessed by measuring EOP value. PP01ΔmotB drastically reduced its infectivity to *E. coli* O157:H7 and confirmed that *motB* was essential for infection to this host (Fig. 4-3). Besides, $T2_{PP01\text{-}g37\text{-}38\text{-}motB}$ displayed infectivity to *E. coli* O157:H7 comparable with PP01, regardless of host for propagation, unlike $T2_{PP01\text{-}g37\text{-}38}$ (Fig. 4-4). Thus it was clarified that gene *37*, *38* and *motB* was responsible for the difference of infectivity between PP01 and T2. It should be also noteworthy that whichever PP01 and $T2_{PP01\text{-}g37\text{-}38}$ were propagated using *E. coli* O157:H7 or *E. coli* C as host, they showed higher titer on *E. coli* C than *E. coli* O157:H7 (Fig. 4-4). It should be important to identify genes of *E. coli* O157 and *E. coli* C which are accounted for their difference of susceptibility for further development of this study.

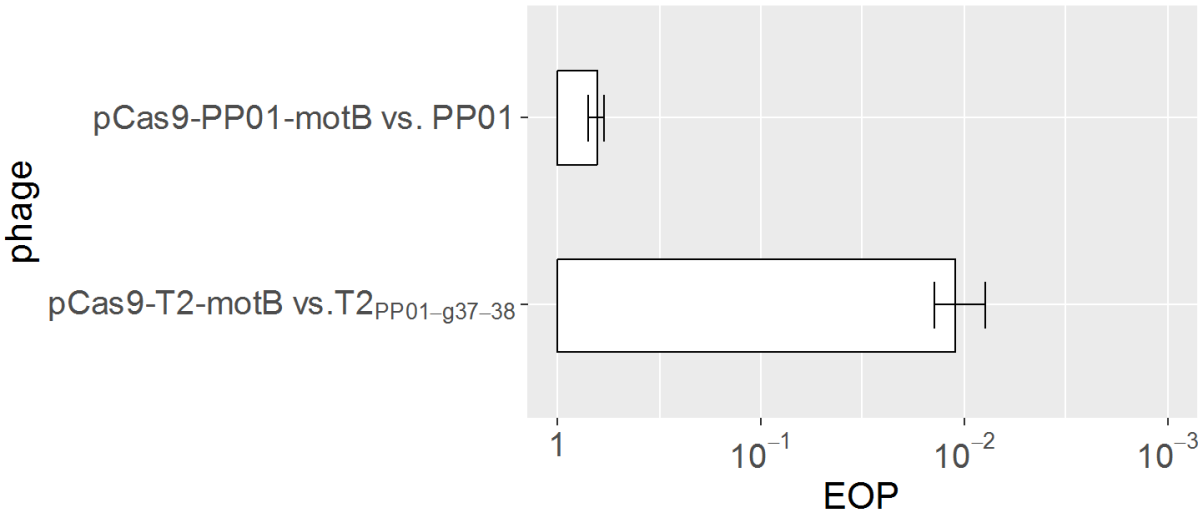


Fig. 4-2 Confirmation of the efficacy of CRISPR/Cas9

EOP was defined as the ratio of phage titer estimated on the host (*E. coli* C) carrying CRISPR/Cas9 to the host without CRISPR/Cas9.

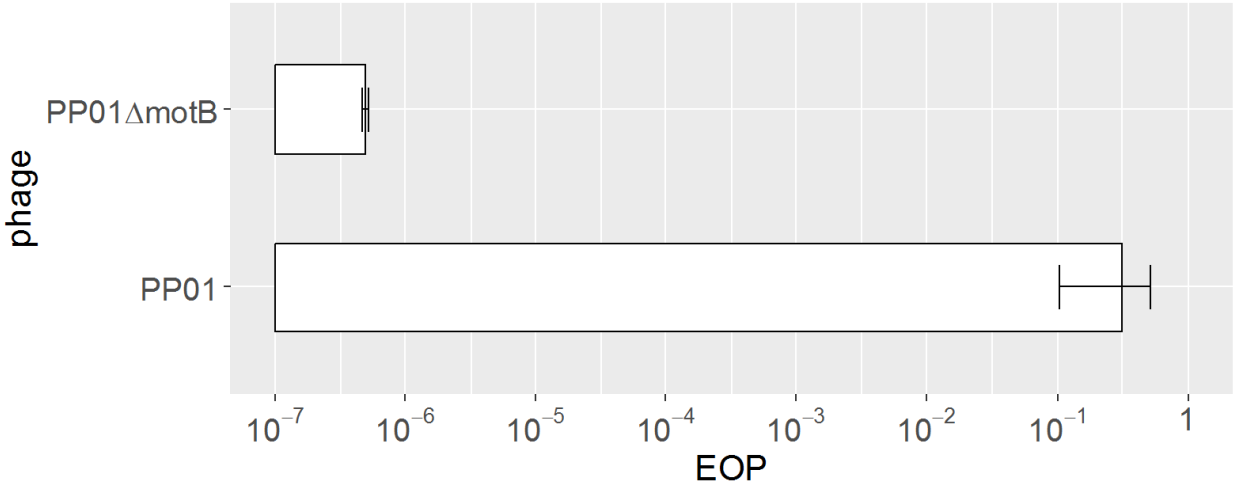


Fig. 4-3 Infectivity of PP01ΔmotB to *E. coli* O157:H7

EOP was defined as the ratio of phage titer estimated on *E. coli* O157:H7 to *E. coli* C. The data represent means±standard deviations of three technical replicates.

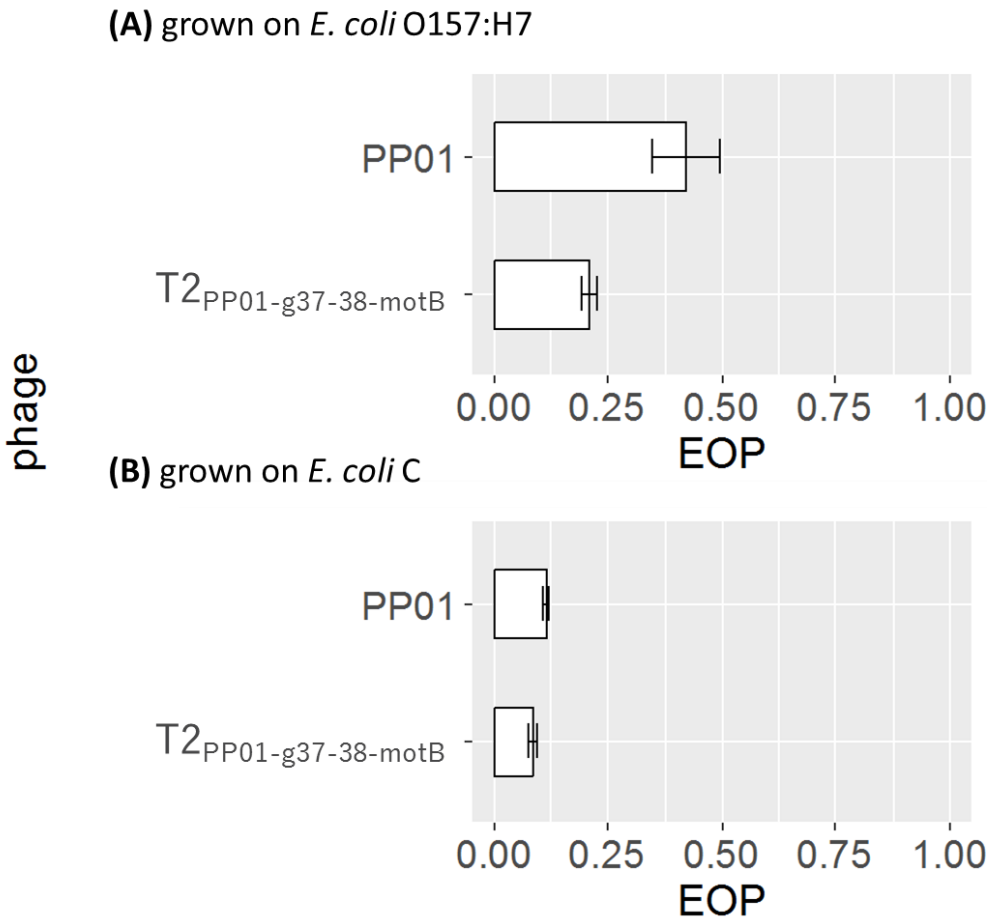


Fig. 4-4 Infectivity of T2$_{PP01\text{-}g37\text{-}38\text{-}motB}$ to *E. coli* O157:H7

EOP of PP01 (control) and T2$_{PP01\text{-}g37\text{-}38\text{-}motB}$ which were propagated using (A) *E. coli* O157:H7 and (B) *E. coli* C as host are displayed. Here EOP was defined as the ratio of phage titer estimated on *E. coli* O157:H7 to *E. coli* C. The data represent means ± standard deviations of three technical replicates.

4.3.2 Phylogenetic analysis of *motB* of T4-like virus

Phylogenetic tree of MotB homologs of 66 T4-like phages including O157:H7 specific phages (AR1, ECML-134, O157-typing phage 7, PP01, vB_EcoM_112, wV7) were displayed in Fig. 4-5. MotB homologs were roughly classified into 4 major clusters. Those of O157:H7 specific phages were dispersed into the 4 clusters and PP01's MotB were distinct from the others. If *motB* is widely important for infection to not only *E. coli* O157:H7 used in this study (ATCC43888) but the other *E. coli* O157:H7 strains, *motB* should be conserved among O157:H7 specific phages. Therefore the importance of PP01's *motB* for infection observed here should be specific to ATCC43888 and regardless of O157:H7 serotype.

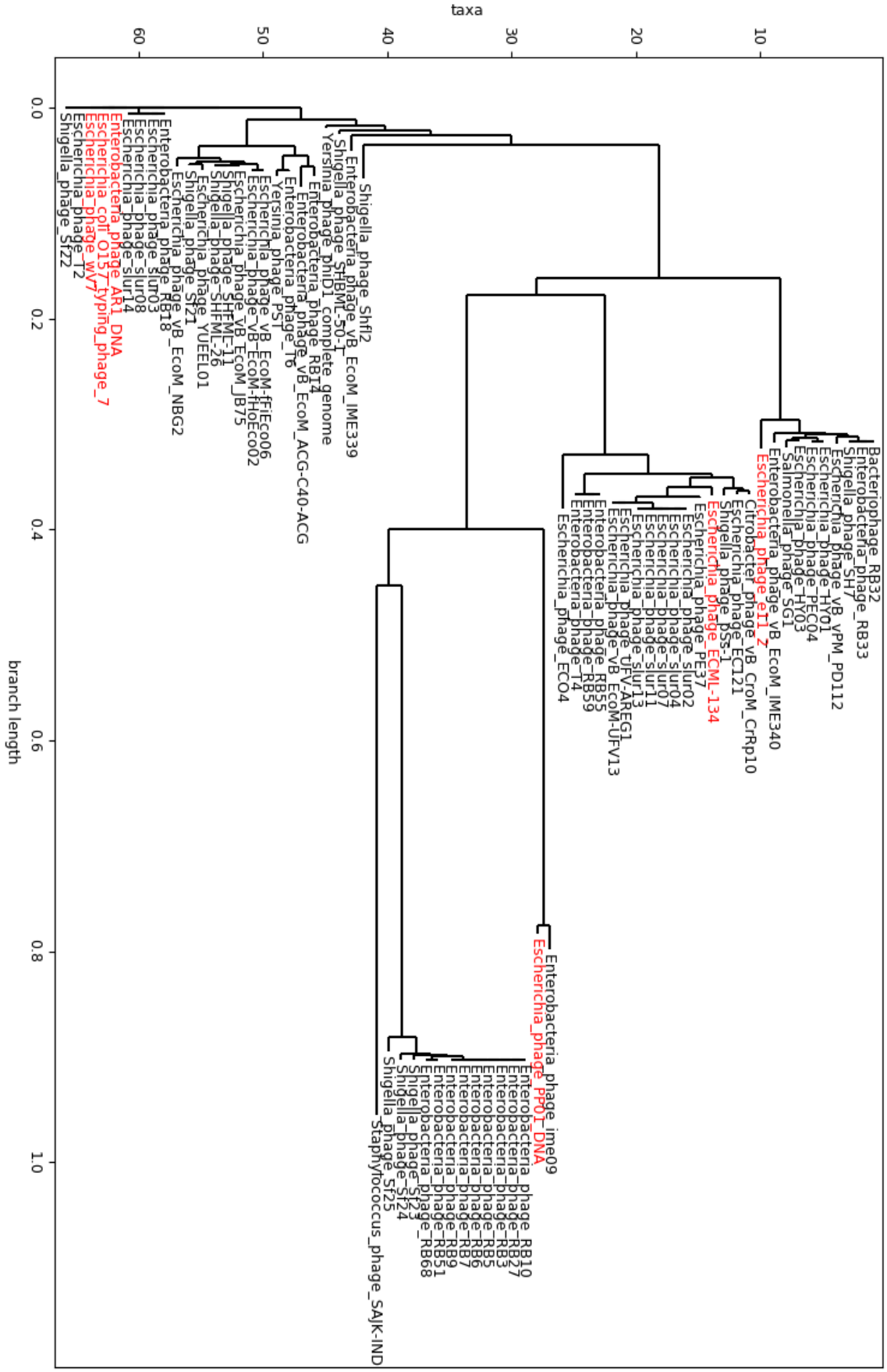

taxa
branch length
Bacteriophage_RB32
Enterobacteria_phage_RB33
Shigella_phage_SH7
Escherichia_phage_vB_vPM_PD112
Escherichia_phage_HY01
Escherichia_phage_PEC04
Escherichia_phage_HY03
Salmonella_phage_SG1
Enterobacteria_phage_vB_EcoM_IME340
Escherichia_phage_e11_2
Citrobacter_phage_vB_CroM_CrRp10
Escherichia_phage_EC121
Shigella_phage_pSs-1
Escherichia_phage_ECML-134
Escherichia_phage_PE37
Escherichia_phage_slur02
Escherichia_phage_slur04
Escherichia_phage_slur07
Escherichia_phage_slur11
Escherichia_phage_slur13
Escherichia_phage_vB_EcoM-UFV13
Enterobacteria_phage_UFV-AREG1
Enterobacteria_phage_RB55
Enterobacteria_phage_RB59
Enterobacteria_phage_T4
Escherichia_phage_ECO4
Enterobacteria_phage_ime09
Escherichia_phage_PP01_DNA
Enterobacteria_phage_RB10
Enterobacteria_phage_RB27
Enterobacteria_phage_RB3
Enterobacteria_phage_RB5
Enterobacteria_phage_RB6
Enterobacteria_phage_RB7
Enterobacteria_phage_RB49
Enterobacteria_phage_RB51
Enterobacteria_phage_RB68
Shigella_phage_Sf23
Shigella_phage_Sf24
Shigella_phage_Sf25
Staphylococcus_phage_SAJK-IND
Shigella_phage_Shf2
Enterobacteria_phage_vB_EcoM_IME339
Shigella_phage_SHBML-50-1
Yersinia_phage_phiD1_complete_genome
Enterobacteria_phage_RB14
Enterobacteria_phage_T6
Enterobacteria_phage_vB_EcoM_ACG-C40-ACG
Yersinia_phage_PST
Escherichia_phage_vB_EcoM-ffEco06
Escherichia_phage_vB_EcoM-fHoEco02
Escherichia_phage_vB_EcoM_JB75
Shigella_phage_SHFML-11
Shigella_phage_SHFML-26
Escherichia_phage_YUEEL01
Shigella_phage_Sf21
Escherichia_phage_vB_EcoM_NBG2
Enterobacteria_phage_RB18
Escherichia_phage_slur03
Escherichia_phage_slur08
Escherichia_phage_slur14
Enterobacteria_phage_AR1_DNA
Escherichia_coli_O157_typing_phage_7
Escherichia_phage_wV7
Escherichia_phage_T2
Shigella_phage_Sf22
10
20
30
40
50
60
0.0
0.2
0.4
0.6
0.8
1.0

Fig 4-5 Phylogenetic tree of *motB* of T4-like phages

The phylogenetic tree was drawn based on the multiple alignment of the amino acid sequences of *motB* from 66 T4-like phages. The letter represents the specie name of each of phages. The names of *E. coli* O157:H7 specific phages were show in red.

4.4 Discussion

As expected from the success of T2 phage, CRISPR/Cas9 could be also applied to genome editing of PP01. This is the first report of genome editing of O157:H7 specific phage using CRISPR/Cas as far as I know. It would be promising that the same method is available for other T4-like O157:H7 specific phage.

PP01ΔmotB, *motB* deletion mutant of PP01 constructed using this method, confirmed that *motB* is the essential gene for PP01 to infect *E. coli* O157:H7. Besides, introducing PP01's *motB* into $T2_{PP01\text{-}g3738}$ could make the recombinant phage ($T2_{PP01\text{-}g37\text{-}38\text{-}motB}$) able to infect *E. coli* O157:H7. Thus it was concluded that the difference of the infectivity to *E. coli* O157:H7 between PP01 and T2 was accounted for the dissimilarity of the long tail fiber and *motB*.

In the previous studies, T4 MotB (the only characterized homolog of PP01 MotB) was considered to optimize the level of T4 middle transcription and named for "modifier of transcription B" [65]. However, the recent report showed that T4 MotB is not related to transcription and seems to counteract with a certain host defense mechanism [66]. *E. coli* histone-like protein, H-NS and StpA is known to attach with foreign DNA and inhibit its transcription [67]. The study revealed that T4 MotB protein can be co-purified with H-NS and StpA in pull-down assay and absence of *motB* decreases the burst size of T4 phage 2-fold. That implies T4 MotB possibly can interact with H-NS and StpA to hinder their

function inhibiting infection of T4, although the exact mechanism still remain to be known [66]. Similarly to T4 MotB, the function of PP01 MotB might be counter against host defense mechanism. But its importance is incomparable with T4 MotB because the absence of PP01 MotB reduced EOP ~10^6 fold. In addition, it is unlikely that PP01 MotB interacts with H-NS and StpA like T4 MotB, since the amino acid sequences of those gene were identical between *E. coli* O157:H7 and *E. coli* C (data not shown). Therefore interaction of PP01 MotB and the unknown host defense mechanism might be novel and quite distinct from one of T4 MotB and H-NS/StpA. Comparison of MotB homologs of T4 like phages implied the presence of this host defense was not specific to *E. coli* O157:H7 strains. If genes encoding this system is identified and its prevalence in *E. coli* genome is revealed, that should be helpful to predict host-phage interaction from their genomic information.

CHAPTER 5 Conclusion and Perspective

Understanding of host-phage interaction is important to exploit phage as an antimicrobial agent. Identification of phage genes important for infection to the host and host genes responsible for phage defense should provide a more reasonable strategy for practical application of phage therapy such as anticipation of phage infectivity from their genomic information and artificial alteration of host range.

In this study, it was attempted to identify the genes responsible for PP01's infectivity specific to *E. coli* O157:H7 based on comparison with non-O157:H7 phage, T2. By whole genome analysis and genome editing using CRISPR/Cas9, it turned out that gene *37* and *38* encoding the distal long tail fiber and gene *motB*, whose function remains to be known, were essential for infection to *E. coli* O157:H7. Besides, introducing those genes into T2 phage could even enable the T2 recombinant to infect *E. coli* O157:H7 (see summary in Fig. 5-1). On top of that, the replacement of the long tail fiber and the short tail fiber tip of T2 with the counterpart of PP01 endow the T2 mutant with adsorption ability comparable with PP01, which is not crucial for phage infectivity but still important parameter related to phage fitness. Thus the results could not only identify the essential gene of PP01's infectivity but develop a proof of concept of artificial host range alteration.

It could be revealed that PP01 MotB play an important role for infection to *E. coli* O157:H7. However, its mechanism of action was little understood in this study. As described in chapter 2, around 10^{-6} of the population of $T2_{PP01g37\text{-}38}$ could infect *E. coli* O157:H7 but those plaques lost the infectivity to this host after grown on *E. coli* C. So they probably obtained a different pattern of genome modification rather than spontaneous mutation through propagation on *E. coli* O157:H7. Although in this report any conclusive evidence couldn't be provided for the relationship between modification

of phage genome and infectivity *E. coli* O157:H7, this point could be an important clue to understand the mechanism of PP01 MotB. Therefore, to further develop this study, it is intriguing to analyze mutations and methylation pattern of the progeny of $T2_{PP01g37\text{-}38}$ which escaped the defense of *E. coli* O157:H7. A new type next generation sequencer can determine DNA sequence and base modification at same time [68]. Such analysis should be also performed using PP01ΔmotB.

Besides, it is also crucial to identify gene(s) encoding that hypothetical defense of *E. coli* O157:H7 interacting with this phage protein. Genome comparison between this bacteria and *E. coli* C should provide useful insight for this issue. It would be also useful to explore genotype of *E. coli* O157:H7 susceptible to PP01ΔmotB by transposon mutagenesis. If such gene(s) is identified and its distribution can be evaluated among other pathogenic *E. coli* strains, that should provide useful information to select appropriate phage for therapeutic application.

In addition to suggesting a more reasonable strategy for phage therapy, the research of host-phage interaction might be beneficial to development of genetic engineering. The study of the phage-defense mechanism of bacteria has been providing a novel tool for gene manipulation such as restriction enzymes and CRISPR/Cas system. So far there is only one report of the interaction of MotB and host defense (T4 MotB and H-NS/SNp1), and the action of PP01 MotB might be related to that but presumably quite distinct. Thus the study of PP01 MotB would be a quite untouched and interesting field. The detailed analysis of the molecular mechanism of PP01 MotB to counteract against host defense might even lead to a discovery of new device of genetic engineering.

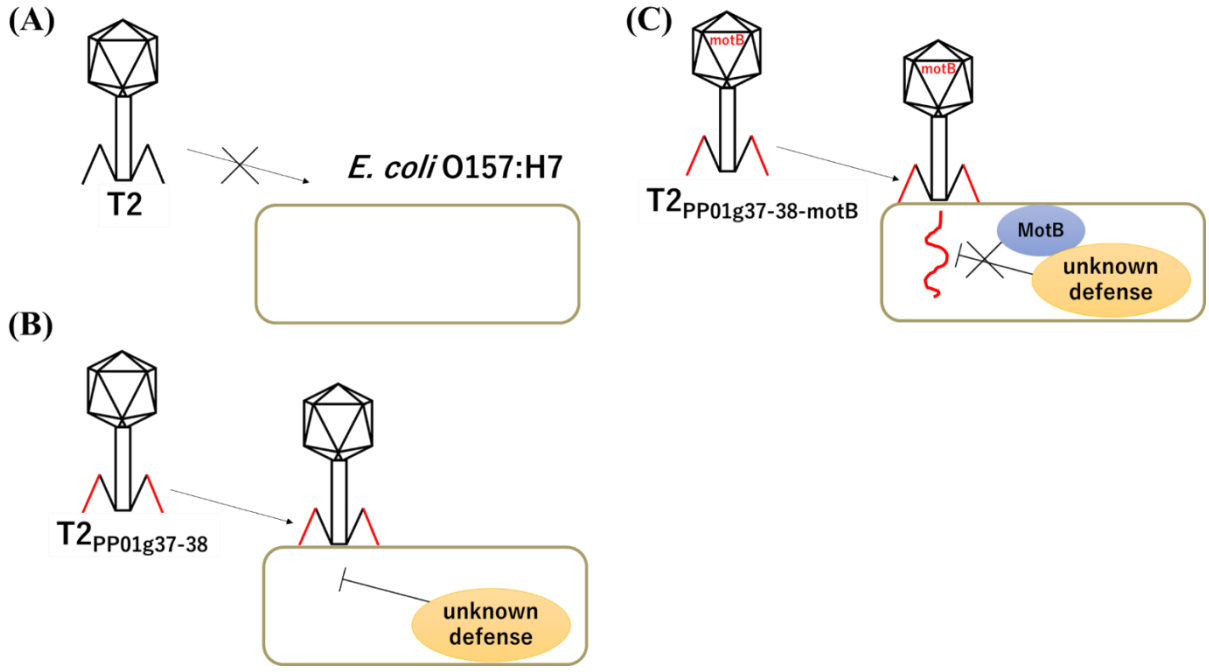


Fig. 5-1 Summary of interaction of T2 phage between *E. coli* O157:H7 and the implied underlying mechanism

T2 phage couldn't attach with the cell surface of *E. coli* O157:H7 (A). $T2_{PP01g37-38}$, the T2 recombinant whose long tail fiber is replaced with one of PP01 (shown in red) could adsorb to the host but the following infection process seemed unable to proceed presumably due to unknown host defense mechanism (B). Further introduction of *motB* enabled the T2 recombinant ($T2_{PP01\text{-}g37\text{-}38\text{-}motB}$) to normally infect *E. coli* O157:H7, implying MotB of PP01 could inactivate the hypothetical blockage (C).

List of Publications

- **F. Hoshiga**, K. Yoshizaki, N. Takao, K. Miyanaga, Y. Tanji., Modification of T2 phage infectivity toward *Escherichia coli* O157:H7 via using CRISPR/Cas9, FEMS Microbiology Letters, doi: 10.1093/femsle/fnz041
- A.H. Azam, **F. Hoshiga**, I. Takeuchi, K. Miyanaga, Y. Tanji. 2018. Analysis of phage-resistant mechanism in *Staphylococcus aureus* SA003 reveals a different binding mechanism for the closely related twort-like phages ϕSA012 and ϕSA039. Appl Microbiol Biotechnol, 2018 Oct;102(20):8963-8977. doi: 10.1007/s00253-018-9269-x.
- T. Tosi, **F. Hoshiga**, C. Millership1, R. Singh, C. Eldrid, D. Patin, D. Mengin-Lecreulx, K. Thalassinos, P. Freemont, A. Gründling. Inhibition of the Staphylococcus aureus c-di-AMP cyclase DacA by direct interaction with the phosphoglucomutase GlmM, PLOS Pathogens, 2019 vol: 15 (1) pp: e1007537. doi: 10.1371/journal.ppat.1007537

Acknowledgement

The present work has been performed under the supervision of Professor Yasunori Tanji and Dr. Kazuhiko Miyanaga in the School of Life Science and Technology, Tokyo Institute of Technology. I sincerely appreciate all of their supports and advises for my research.

I am grateful to all students and faculties of the Academy of Computational Life Science, which gave me a great opportunity to expand my skill. I would also thank the ministry of education, culture, sport, and technology (MEXT) of Japan for financial support through ACLS.

I wish to thank all members of Prof. Tanji laboratory for their corporation, especially for the collaboration work with Mr. Aa Haeruman Azam.

Last but not least, I would like to thank my parents for all kind support and encouragement for my study.

Reference

[1] C. Loc-Carrillo and S. T. Abedon, “Pros and cons of phage therapy,” *Bacteriophage*, vol. 1, no. 2, pp. 111–114, 2011.

[2] G. P. C. Salmond and P. C. Fineran, “A century of the phage: past, present and future.,” *Nat. Rev. Microbiol.*, vol. 13, no. 12, pp. 777–86, 2015.

[3] J. Gu *et al.*, “A method for generation phage cocktail with great therapeutic potential,” *PLoS One*, vol. 7, no. 3, 2012.

[4] A. Oliveira, R. Sereno, and J. Azeredo, “In vivo efficiency evaluation of a phage cocktail in controlling severe colibacillosis in confined conditions and experimental poultry houses,” *Vet. Microbiol.*, vol. 146, no. 3–4, pp. 303–308, 2010.

[5] Y. Tanji, T. Shimada, H. Fukudomi, K. Miyanaga, Y. Nakai, and H. Unno, “Therapeutic use of phage cocktail for controlling Escherichia coli O157:H7 in gastrointestinal tract of mice,” *J. Biosci. Bioeng.*, vol. 100, no. 3, pp. 280–287, 2005.

[6] A. Wright, C. H. Hawkins, E. E. Änggård, and D. R. Harper, “A controlled clinical trial of a therapeutic bacteriophage preparation in chronic otitis due to antibiotic-resistant Pseudomonas aeruginosa; A preliminary report of efficacy,” *Clin. Otolaryngol.*, vol. 34, no. 4, pp. 349–357, 2009.

[7] T. K. Lu and M. S. Koeris, “The next generation of bacteriophage therapy,” *Curr. Opin. Microbiol.*, vol. 14, no. 5, pp. 524–531, 2011.

[8] S. J. Labrie, J. E. Samson, and S. Moineau, “Bacteriophage resistance mechanisms,” *Nat. Rev. Microbiol.*, vol. 8, no. 5, pp. 317–327, 2010.

[9] J. E. Samson, A. H. Magadán, M. Sabri, and S. Moineau, “Revenge of the

phages: defeating bacterial defences.," *Nat. Rev. Microbiol.*, vol. 11, no. 10, pp. 675–87, 2013.

[10] W.-C. Liao, W. V. Ng, I.-H. Lin, W.-J. Syu, T.-T. Liu, and C.-H. Chang, "T4-Like genome organization of the Escherichia coli O157:H7 lytic phage AR1.," *J. Virol.*, vol. 85, no. 13, pp. 6567–6578, 2011.

[11] T. Abuladze, M. Li, M. Y. Menetrez, T. Dean, A. Senecal, and A. Sulakvelidze, "Bacteriophages reduce experimental contamination of hard surfaces, tomato, spinach, broccoli, and ground beef by Escherichia coli O157:H7," *Appl. Environ. Microbiol.*, vol. 74, no. 20, pp. 6230–6238, 2008.

[12] L. A. Cowley *et al.*, "Analysis of whole genome sequencing for the Escherichia coli O157: H7 typing phages," *BMC Genomics*, vol. 16, no. 1, pp. 1–13, 2015.

[13] G. O. Flynn, R. P. Ross, G. F. Fitzgerald, and A. Coffey, "Evaluation of a Cocktail of Three Bacteriophages for Biocontrol of Escherichia coli O157 : H7 Evaluation of a Cocktail of Three Bacteriophages for Biocontrol of Escherichia coli O157 : H7," *Appl. Environ. Microbiol.*, vol. 70, no. 6, pp. 3417–3424, 2004.

[14] a. M. Kropinski *et al.*, "Escherichia coli O157:H7 Typing Phage V7 Is a T4-Like Virus," *J. Virol.*, vol. 86, no. 18, pp. 10246–10246, 2012.

[15] C. L. Gyles, "Shiga toxin-producing Escherichia coli: an overview.," *J. Anim. Sci.*, vol. 85, no. 13 Suppl, 2007.

[16] G. Z. Panos, G. I. Betsi, and M. E. Falagas, "Systematic review: Are antibiotics detrimental or beneficial for the treatment of patients with Escherichia coli O157:H7 infection?," *Aliment. Pharmacol. Ther.*, vol. 24, no. 5, pp. 731–742, 2006.

[17] S. Sabouri, Z. Sepehrizadeh, S. Amirpour-Rostami, and M. Skurnik, "A

minireview on the in vitro and in vivo experiments with anti-Escherichia coli O157:H7 phages as potential biocontrol and phage therapy agents," *Int. J. Food Microbiol.*, vol. 243, pp. 52–57, 2017.

[18] F. Yu and S. Mizushima, "Roles of lipopolysaccharide and outer membrane protein ompc of Escherichia coli K-12 in the receptor function for bacteriophage T4," *J. Bacteriol.*, vol. 151, no. 2, pp. 718–722, 1982.

[19] I. Riede, "Receptor specificity of the short tail fibres (gp12) of T-even type Escherichia coli phages," *MGG Mol. Gen. Genet.*, vol. 206, no. 1, pp. 110–115, 1987.

[20] E. S. Miller *et al.*, "Bacteriophage T4 Genome Bacteriophage T4 Genome †," *Microbiol. Mol. Biol. Rev.*, vol. 67, no. 1, pp. 86–156, 2003.

[21] M. Yoichi, M. Abe, K. Miyanaga, H. Unno, and Y. Tanji, "Alteration of tail fiber protein gp38 enables T2 phage to infect Escherichia coli O157:H7," *J. Biotechnol.*, vol. 115, no. 1, pp. 101–107, 2005.

[22] C. Agbemabiese and A. Singh, "Optimal adsorption rate : Implications of the shielding effect," pp. 2140–2145, 2017.

[23] M. Delbruck, "Adsorption of Bacteriophage Under Various Physiological Conditions of the Host," *J. Gen. Physiol.*, vol. 23, no. 5, pp. 631–642, 1940.

[24] R. Gallet, Y. Shao, and I.-N. Wang, "High adsorption rate is detrimental to bacteriophage fitness in a biofilm-like environment," *BMC Evol. Biol.*, vol. 9, no. 1, p. 241, 2009.

[25] Y. Shao and I. N. Wang, "Bacteriophage adsorption rate and optimal lysis time," *Genetics*, vol. 180, no. 1, pp. 471–482, 2008.

[26] W. Arber, "HOST-CONTROLLED MODIFICATION OF BACTERIOPHAGE,"

Annu. Rev. Microbiol., pp. 365–378, 1964.

[27] J. Casadesús and R. D'Ari, "Memory in bacteria and phage," *BioEssays*, vol. 24, no. 6, pp. 512–518, 2002.

[28] H. E. Selick, K. N. Kreuzer, and B. M. Alberts, "The bacteriophage T4 insertion/substitution vector system. A method for introducing site-specific mutations into the virus chromosome," *J. Biol. Chem.*, vol. 263, no. 23, pp. 11336–11347, 1988.

[29] A. M. Box, M. J. McGuffie, B. J. O'Hara, and K. D. Seed, "Functional analysis of bacteriophage immunity through a Type I-E CRISPR-Cas system in Vibrio cholerae and its application in bacteriophage genome engineering," *J. Bacteriol.*, vol. 198, no. 3, pp. 578–590, 2015.

[30] R. Kiro, D. Shitrit, and U. Qimron, "Efficient engineering of a bacteriophage genome using the type I-E CRISPR-Cas system.," *RNA Biol.*, vol. 11, no. 1, pp. 42–4, 2014.

[31] M.-L. Lemay, D. M. Tremblay, and S. Moineau, "Genome Engineering of Virulent Lactococcal Phages Using CRISPR-Cas9," *ACS Synth. Biol.*, vol. 6, no. 7, pp. 1351–1358, Jul. 2017.

[32] B. Martel and S. Moineau, "CRISPR-Cas: An efficient tool for genome engineering of virulent bacteriophages," *Nucleic Acids Res.*, vol. 42, no. 14, pp. 9504–9513, 2014.

[33] P. TAO, X. Wu, W. Tang, J. Zhu, and V. Rao, "Engineering of Bacteriophage T4 Genome Using CRISPR-Cas9," *ACS Synth. Biol.*, p. acssynbio.7b00179, 2017.

[34] A. S. Of, H. Between, T. W. O. Strains, and O. F. Escherichia, "A STUDY OF HYBRIDS BETWEEN TWO STRAINS OF ESCHERICHIA COLI," vol. 1, pp.

468–471, 1954.

[35] H. Wilson, S. Datta, and a Oppenheim, “SPECIALIZED TECHNIQUES UNIT 1.16 Recombineering: Genetic Engineering in Bacteria Using Homologous Recombination,” *Curr. Protoc. Mol. Biol.*, vol. 1, p. 1, 2007.

[36] W. Jiang, D. Bikard, D. Cox, F. Zhang, and L. a Marraffini, “RNA-guided editing of bacterial genomes using CRISPR-Cas systems,” *Nat Biotechnol*, vol. 31, no. 3, pp. 233–239, 2013.

[37] M. J. van Raaij, G. Schoehn, M. R. Burda, and S. Miller, “Crystal structure of a heat and protease-stable part of the bacteriophage T4 short tail fibre,” *J. Mol. Biol.*, vol. 314, no. 5, pp. 1137–1146, 2001.

[38] P. Rice, I. Longden, and A. Bleasby, “EMBOSS: the European Molecular Biology Open Software Suite.,” *Trends Genet.*, vol. 16, no. 6, pp. 276–7, Jun. 2000.

[39] R Core Team, “R: A Language and Environment for Statistical Computing.” Vienna, Austria, 2018.

[40] V. M. Petrov, S. Ratnayaka, J. M. Nolan, E. S. Miller, and J. D. Karam, “Genomes of the T4-related bacteriophages as windows on microbial genome evolution,” *Virol. J.*, vol. 7, no. 1, p. 292, 2010.

[41] R. Revel, “Restriction of Nonglucosylated Permissive Mutants Bacteriophage : co / i B and Properties of of Escherichia,” vol. 701, pp. 688–701, 1967.

[42] G. G. Wilson, V. I. Tanyashin, and N. E. Murray, “Molecular cloning of fragments of bacteriophage T4 DNA.,” *Mol. Gen. Genet.*, vol. 156, no. 2, pp. 203–14, Nov. 1977.

[43] V. Tcherepanov, A. Ehlers, and C. Upton, “Genome Annotation Transfer Utility

(GATU): rapid annotation of viral genomes using a closely related reference genome.," *BMC Genomics*, vol. 7, p. 150, 2006.

[44] J. Besemer and M. Borodovsky, "GeneMark: web software for gene finding in prokaryotes, eukaryotes and viruses.," *Nucleic Acids Res.*, vol. 33, no. Web Server issue, pp. W451-4, Jul. 2005.

[45] T. M. Lowe and P. P. Chan, "tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes," *Nucleic Acids Res.*, vol. 44, no. W1, pp. W54–W57, 2016.

[46] A. E. Darling, B. Mau, and N. T. Perna, "Progressivemauve: Multiple genome alignment with gene gain, loss and rearrangement," *PLoS One*, vol. 5, no. 6, 2010.

[47] M. et al Krzywinski, "Circos: an Information Aesthetic for Comparative Genomics," *Genome Res*, vol. 19, no. 604, pp. 1639–1645, 2009.

[48] G. Moreno-hagelsieb and K. Latimer, "Sequence analysis Choosing BLAST options for better detection of orthologs as reciprocal best hits," vol. 24, no. 3, pp. 319–324, 2017.

[49] S. F. Altschup, W. Gish, T. Pennsylvania, and U. Park, "Basic Local Alignment Search Tool 2Department of Computer Science," pp. 403–410, 1990.

[50] J. Brooks and S. Hattman, "Location of the DNA-adenine methylase gene on the genetic map of phage T2," *Virology*, vol. 55, no. 1, pp. 285–288, 1973.

[51] W. Xu, P. Gauss, J. Shen, C. A. Dunn, and M. J. Bessman, "The gene e.1 (nudE.1) of T4 bacteriophage designates a new member of the nudix hydrolase superfamily active on flavin adenine dinucleotide, adenosine 5′-triphospho-5′-adenosine, and ADP-ribose," *J. Biol. Chem.*, vol. 277, no. 26, pp. 23181–23185,

2002.

[52] D. R. Edgell, E. A. Gibb, and M. Belfort, “Mobile DNA elements in T4 and related phages,” *Virol. J.*, vol. 7, pp. 1–15, 2010.

[53] S. Chibani-chennoufi, C. Canchaya, A. Bruttin, H. Brüssow, and H. Bru, “Comparative Genomics of the T4-Like Escherichia coli Phage JS98 : Implications for the Evolution of T4 Phages Comparative Genomics of the T4-Like Escherichia coli Phage JS98 : Implications for the Evolution of T4 Phages,” *J. Bacteriol.*, vol. 186, no. 24, pp. 8276–8286, 2004.

[54] A. M. Comeau, C. Bertrand, A. Letarov, F. Tétart, and H. M. Krisch, “Modular architecture of the T4 phage superfamily: A conserved core genome and a plastic periphery,” *Virology*, vol. 362, no. 2, pp. 384–396, 2007.

[55] J. M. Mullaney and L. W. Black, “Capsid targeting sequence targets foreign proteins into bacteriophage T4 and permits proteolytic processing,” *J. Mol. Biol.*, vol. 261, no. 3, pp. 372–385, 1996.

[56] D. Rifat, N. T. Wright, K. M. Varney, D. J. Weber, and L. W. Black, “Restriction Endonuclease Inhibitor IPI* of Bacteriophage T4: A Novel Structure for a Dedicated Target,” *J. Mol. Biol.*, vol. 375, no. 3, pp. 720–734, 2008.

[57] H. Deveau, J. E. Garneau, and S. Moineau, “CRISPR/Cas System and Its Role in Phage-Bacteria Interactions,” *Annu. Rev. Microbiol.*, vol. 64, no. 1, pp. 475–493, Oct. 2010.

[58] W. B. Wood and M. P. Conley, “Attachment of tail fibers in bacteriophage T4 assembly: Role of the phage whiskers,” *J. Mol. Biol.*, vol. 127, no. 1, pp. 15–29, 1979.

[59] T. Ishii and M. Yanagida, “The two dispensable structural proteins (soc and hoc)

of the T4 phage capsid; their purification and properties, isolation and characterization of the defective mutants, and their binding with the defective heads in vitro.," *J. Mol. Biol.*, vol. 109, no. 4, pp. 487–514, Feb. 1977.

[60] A. Fokine, M. Z. Islam, Z. Zhang, V. D. Bowman, V. B. Rao, and M. G. Rossmann, "Structure of the Three N-Terminal Immunoglobulin Domains of the Highly Immunogenic Outer Capsid Protein from a T4-Like Bacteriophage," *J. Virol.*, vol. 85, no. 16, pp. 8141–8148, 2011.

[61] M. Yanagida, Y. Suzuki, and T. Toda, "Molecular organization of the head of bacteriophage Teven: underlying design principles.," *Adv. Biophys.*, vol. 17, pp. 97–146, 1984.

[62] E. W. Sayers *et al.*, "Database resources of the National Center for Biotechnology Information," *Nucleic Acids Res.*, vol. 37, no. Database, pp. D5–D15, Jan. 2009.

[63] M. A. Larkin *et al.*, "Clustal W and Clustal X version 2.0," *Bioinformatics*, vol. 23, no. 21, pp. 2947–2948, Nov. 2007.

[64] P. J. A. Cock *et al.*, "Biopython: freely available Python tools for computational molecular biology and bioinformatics," *Bioinformatics*, vol. 25, no. 11, pp. 1422–1423, Jun. 2009.

[65] J. F. Pulitzer, M. Colombot, and M. Ciaramella, "New Control Elements of Bacteriophage Pre-replicative Transcription T4," pp. 249–263, 1985.

[66] J. Patterson-West *et al.*, "The bacteriophage T4 MotB protein, a DNA-binding protein, improves phage fitness," *Viruses*, vol. 10, no. 7, 2018.

[67] D. C. Grainger, "Structure and function of bacterial H-NS protein," *Biochem. Soc. Trans.*, vol. 44, no. 6, pp. 1561–1569, Dec. 2016.

[68] Y. Zhang and A. Jeltsch, “The Application of Next Generation Sequencing in DNA Methylation Analysis,” pp. 85–101, 2010.