

Isolation and characterization of a novel temperate *Escherichia coli* bacteriophage, Kapi1, which modifies the O-antigen and contributes to the competitiveness of its host during colonization of the murine gastrointestinal tract.

Kat Pick^a, Tingting Ju^b, Benjamin P. Willing^b, and Tracy Raivio^{a,#}

^a Department of Biological Sciences, University of Alberta, Edmonton, Alberta, Canada

^b Department of Agricultural, Food and Nutritional Science, University of Alberta, Edmonton, Alberta, Canada

Running Title: Novel temperate phage contributes to host fitness

Address correspondence to Tracy Raivio, traivio@ualberta.ca

Abstract word count: 177 (Importance: 121)

Text word count: 6320

Abstract

In this study, we describe the isolation and characterization of novel bacteriophage vB_EcoP_Kapi1 (Kapi1) isolated from a strain of commensal *Escherichia coli* inhabiting the gastrointestinal tract of healthy mice. We show that Kapi1 is a temperate phage integrated into tRNA *argW* of strain MP1 and describe its genome annotation and structure. Kapi1 shows limited homology to other characterized prophages but is most similar to the seroconverting phages of *Shigella flexneri*, and clusters taxonomically with P22-like phages. The receptor for Kapi1 is the lipopolysaccharide O-antigen, and we further show that Kapi1 alters the structure of its hosts O-antigen in multiple ways. Kapi1 displays unstable lysogeny, and we find that lysogeny is favored during growth in simulated intestinal fluid. Furthermore, Kapi1 lysogens have a competitive advantage over their non-lysogenic counterparts both *in vitro* and *in vivo*, suggesting a role for Kapi1 during colonization. We thus report the use of MP1 and Kapi1 as a model system to explore the molecular mechanisms of mammalian colonization by *E. coli* to ask what the role(s) of prophages in this context might be.

Importance

Although research exploring the microbiome has exploded in recent years, our understanding of the viral component of the microbiome is lagging far behind our understanding of the bacterial component. The vast majority of intestinal bacteria carry prophages integrated into their chromosomes, but most of these bacteriophages remain uncharacterized and unexplored. Here, we isolate and characterize a novel temperate bacteriophage infecting a commensal strain of *Escherichia coli*. We aim to explore the interactions between

bacteriophages and their hosts in the context of the gastrointestinal tract, asking what role(s) temperate bacteriophage may play in growth and survival of bacteria in the gut. Understanding the fundamental biology of gut commensal bacteria can inform the development of novel antimicrobial or probiotic strategies for intestinal infections.

Introduction

Escherichia coli is a Gram-negative bacterium normally inhabiting the lower gastrointestinal (GI) tract of humans and other warm-blooded animals (1). Despite being one of the most widely studied prokaryotic model organisms, there remains an immense complexity to the lifestyle of *E. coli* that we are only beginning to appreciate; one of these layers of complexity is the interactions between *E. coli* and the bacteriophages that infect it. Bacteriophage (or simply phage) exhibit two main lifestyles; lytic phage infect and immediately begin replicating within their host, eventually causing cell lysis and releasing progeny phages. Temperate phage replicate through the same lytic cycle, but can also display an alternate life cycle, the lysogenic cycle, where the phage integrates into the genome of their host and resides as a mostly-dormant prophage, replicating along with the host chromosome and being disseminated into daughter cells. Once the host cell experiences stress such as DNA damage, the prophage excises from the chromosome and enters the lytic cycle to ensure its own survival (2). Temperate phage have been gaining attention as we begin to appreciate their abundance; it has been estimated that approximately half of all sequenced bacterial genomes contain intact prophage, and even more contain prophage elements (3). Interestingly, the abundance of temperate phage residing in the commensal gut microbiome of mice appears to be even higher (4), indicating that temperate phage may play a role in bacterial community dynamics during colonization. Indeed, many

recent studies and reviews have highlighted the importance of phages in the microbiome community (5–8).

One of the ways in which temperate phage influence the biology of their hosts is through lysogenic conversion. During lysogenic conversion, accessory genes encoded on the prophage are expressed in the host cell during lysogeny. These accessory genes influence the biology of the host cell, without affecting the phage life cycle. One form of lysogenic conversion is seroconversion, in which bacteriophages encode proteins that alter the structure of the host lipopolysaccharide (LPS) O-antigen. The most well-known seroconverting phage are those that infect *Shigella flexneri*; lysogeny with these phage results in modification of the O-antigen through either glucosylation or O-acetylation, leading to a change in bacterial serotype (9). This can have different benefits for the bacterial host including antigenic variation and immune evasion, since the mammalian innate immune system mounts a serotype-specific antibody response (9). Beyond immunogenicity, LPS is an essential component of the outer membrane that is important for membrane stability and barrier function (10).

Here, we describe the isolation and characterization of novel bacteriophage Kapi1, capable of O-antigen modification. Kapi1 was isolated from a wild commensal strain of *E. coli*, and the phage genome was sequenced and compared to other characterized phages. We also report the identification of the O-antigen as the receptor for Kapi1 and show that this phage displays an unstable temperate lifestyle. Our characterization of Kapi1 suggests that it has a significant impact on the fitness of *E. coli* in the GI environment, and that it should serve as an excellent model system to explore the impact of temperate bacteriophage on *E. coli* colonization of the mammalian GI tract.

Results and Discussion

Kapi1 is a novel *Podoviridae* with a narrow host range

Recently, Lasaro et al. (2014) showed that the Cpx, Arc, and Rcs two-component systems found in *Escherichia coli* were required for colonization of the murine GI tract by a strain of commensal *E. coli*, MP1. We began performing competitions between Cpx, Arc, and Rcs mutants and wild-type (WT) MP1 *in vitro* to further explore the molecular mechanisms behind the observed colonization phenotypes. MP1, MP7, and MP13 are identical strains except for the presence of fluorescent plasmids pML8 and pAS07 integrated into the chromosomes of MP7 and MP13 respectively, at the λ attachment site (11). Because MP7 and MP13 are marked with *mcherry* and *gfpmut3.1*, these strains are easily distinguishable during co-culture competition experiments. Unexpectedly, when co-culturing MP13 *rcsB* mutants with MP7, we found that the *rcsB* mutants strongly outcompeted the WT (data not shown). Because Lasaro et al. (2014) showed that mutation of *rcsB* decreased competitiveness in a mouse colonization model, we wondered if this reflected a differential ability of the Rcs mutant to compete *in vitro* vs *in vivo* and set out to investigate this. Because of the strong competitive advantage, we hypothesized that the *rcsB* mutant could perhaps be directly killing the WT in some way. To test if there was a bactericidal factor secreted by the *rcsB* mutant, we isolated the supernatant from cultures of MP13 *rcsB* mutants and spotted it onto lawns of MP7. Unexpectedly, the supernatant cleared the MP7 lawn, and serial dilutions of the supernatant showed spotty clearing, reminiscent of phage plaques. We then screened our entire strain collection of all strains derived from MP1, MP7, and MP13 for their abilities to produce plaques on each other. A clear trend emerged; the supernatants of MP1 and MP13 background strains could produce plaques on lawns of MP7, but

supernatant derived from MP7 background strains could not produce plaques on either MP1 or MP13. Thus, we began identification and characterization of the phage found in MP1 and MP13.

Because our cultures of MP1 and MP13 containing phage did not appear to have a pronounced growth defect in comparison to MP7, we hypothesized that the phage in these cultures may be temperate, as a lytic phage would be more likely to lyse the cultures resulting in a visible reduction in cell density and poor growth. Analysis of the previously published genome sequence for MP1 (11) using the prophage identification tool PHASTER (12, 13) revealed six putative prophages integrated into the chromosome of MP1 (Figure 1A). Of these, only one prophage was scored as intact by PHASTER (12, 13) (Figure 1A); we hypothesized that this prophage was the most likely candidate for the phage plaques we had observed because of the completeness of the prophage sequence. To confirm this, we performed polymerase chain reaction (PCR) on colonies of MP1, MP7 and MP13, as well as on phage lysates prepared from MP1 and MP13 with three primer pairs targeting the coat, portal, and tail proteins within the Sf101-like Intact_1 prophage region identified by PHASTER. Bands were consistently observed for MP1 and MP13 colonies and phage lysates and were consistently absent for MP7 colonies for all primer pairs (data not shown). This indicates that the phage present in MP1 and MP13 cultures likely corresponds to the Sf101-like Intact_1 prophage region identified by PHASTER in the MP1 genome. Despite several attempts to induce phage from MP7 using DNA damaging agents, phage could never be isolated from MP7. Since Lasaro et al. (2014) previously showed that MP7 and MP13 had equal competitive indices *in vivo*, we decided to investigate this further. Upon testing the original stock of MP7 isolated by Lasaro et al. (2014), we found that this strain does in fact contain the Intact_1 prophage region (data not shown), and that only our stock of MP7 lacks the Intact_1 prophage region. These findings are important as they demonstrate that

any competitive advantage that would have been provided by carrying the Intact_1 prophage did not play a role in the findings of Lasaro et al. (2014), since both MP7 and MP13 contain the prophage. We therefore renamed our stock of MP7 which lacks the Intact_1 prophage region to KP7, in order to avoid confusion with the original MP7 stock which does contain the Intact_1 prophage region.

Transmission electron microscopy (TEM) (Figure 1B) of phage lysates collected from MP1 and MP13 revealed phage particles with a mean capsid diameter of 70.10 ± 2.92 nm and tail length 15.37 ± 1.45 nm placing this phage in the family *Podoviridae* and order *Caudovirale*. We screened 11 strains of *E. coli* (Top10, MG1655, TJ-LM, TJ-WM, TJ-LR, MC4100, W3110, BW25113, J96, E2348/69, Nissile 1917) and *Citrobacter rodentium* DBS100 for susceptibility to the phage (see Table S1). Top10, MG1655, MC4100, W3110, and BW25113 are *E. coli* K-12 derivatives (14) whereas TJ-LM, TJ-WM, TJ-LR, J96, E2348/69, and Nissile 1917 are natural *E. coli* isolates (15–18). All strains were completely resistant to infection; from our strain collection, KP7 is the only strain that this phage can infect. PHASTER analysis of each of the strains tested shows no predicted Sf101-like prophage regions; it is therefore unlikely that they are protected from infection via superinfection immunity, but we cannot rule this out entirely as all strains tested are lysogenized by several prophages (data not shown). This preliminary analysis suggests that the newly isolated phage has a relatively narrow host range; a trend which has been observed in other temperate phages isolated from the gut (19). This phage forms diffuse plaques on KP7 with an average plaque diameter of 2.0 ± 0.22 mm after overnight incubation at 37 °C. Although the morphology of phage particles on TEM and plaques on soft-agar overlays were consistent, to confirm that the Intact_1 prophage region is the only prophage in MP1 capable of active excision and lytic replication we performed PCR on DNase-treated phage

lysates using primers corresponding to each putative prophage region identified by PHASTER and a *nuoA* primer pair to control for genomic DNA contamination. No bands were observed in KP7 lysates, and only the band corresponding to Intact_1 prophage was observed in the MP13 lysate (Figure S1).

Kapi1 lacks sequence homology with other *Podoviridae*, and has a modular lambdoid genome

Although the whole genome sequence for MP1 has already been published (11), to be thorough and ensure that the isolated phage was truly a prophage and not introduced by contamination from our laboratory, whole genome sequencing was performed on our stocks of MP1, KP7, and MP13. We also aimed to find the integration site and characterize the genome of the isolated phage. As anticipated, the 39 kb Sf101-like region was present in the genomes of MP1 and MP13, and absent from the genome of KP7. Unfortunately, the phage genome was assembled into its own linear contig, not showing where it may be integrated into the host chromosome. This was observed in both the original MP1 sequence (11) and in our resequencing attempt; it is likely that the phage genome was in its circular form (ie. excised from the host chromosome) since we extracted DNA from late stationary phase cultures. Upon closer analysis of the original MP1 sequence (11), it appears that the ends of contig NZ_JEMI01000030, corresponding to the Intact_1 prophage region, are actually terminal repeats indicating that the sequence is circular. To confirm this circularity, primers were designed pointing outward from each end of the phage contig (prophage_left and prophage_right primers, Table S2), and PCR and Sanger sequencing was performed on DNA extracted from phage lysates. The sequence of the PCR product obtained was consistent with the conclusion that the phage exists in a circular

form at some point during its lifecycle and confirmed complete sequencing of the entire phage genome.

When the phage genome was analysed using BLASTn (20) with the viral filter (taxid: 10239), the top hit was Enterobacteria phage Sf101 (accession NC_027398.1) with 96.47% identity but only 33% query cover, indicating that this phage represents a novel viral species, with less than 95% nucleotide sequence similarity to any other characterized phage (21). We thus named this novel phage vB_EcoP_Kap1 (Kap1, NCBI:txid:2746235). When the viral filter is removed, the top hit still only shows 96.92% identity and 44% query cover to Kap1 (*E. coli* genome assembly FHI87, scaffold-10_contig-14.0_1_42036, accession LM996987.1). Visualization of the alignments between Kap1 and Sf101 (Figure 2) showed that a ~10 kb region of Kap1 corresponding to the virion morphogenesis module is the most conserved region. Further, Kap1 may represent a novel genera of the family *Podoviridae*, since it shares less than 50% nucleotide sequence similarity to any other characterized *Podoviridae* genus (21). Comparing the genome sequence of Kap1 against the type-species for each *Podoviridae* genus in the ICTV 2019.v1 Master Species List (22), the top hit, Enterobacteria phage P22 (accession NC_002371.2), belonging to the genus *Lederbergvirus*, shares only 83.62% identity and 20% query coverage with Kap1. The taxonomy of Kap1 was further explored using vContact2 (23), this analysis showed that Kap1 belongs to the same viral cluster as phages P22 (24) and Sf101 (25) (Figure S2).

The genome of Kap1 (accession MT813197) is 39,436 bp in length and represents 0.83% of the genome of MP1. The GC content of Kap1 is 47.1%, slightly lower than the 50.6% of the host genome. Kap1 has a modular genome structure typical of many lambdoid phages (Figure 2) (26). Beginning from *xis*, the first region of the Kap1 genome is rich in hypothetical

proteins and proteins with unknown function. The next segment of the genome is characterized by the DNA replication/repair/regulation module; this region has a Lambda-like organization, with CIII, N, CI, cro, CII, O, P, and Q. This module is followed by two tRNAs immediately preceding the lysis module (holin, lysin, Rz). The next module is responsible for virion morphogenesis, encoding proteins responsible for the head assembly (terminases, scaffold, portal, coat), followed by tail assembly (DNA stabilization protein, tail needle knob, DNA transfer and ejection proteins). The final module is required for integration, including *xis*, *int*, and *attP*. For a detailed view of annotation and functional assignments for all protein coding sequences (CDS) in Kapi1, see Table S3.

Kapi1 integrates into the 3' end of host tRNA *argW*

To begin our search for the integration site of Kapi1, we used BLASTn (20) to look for prophages similar to Kapi1 in the NCBI nucleotide database. We then analyzed the host chromosome surrounding these Kapi1-like prophages to find any similarities with the MP1 chromosome. We identified two contigs in MP1 whose ends shared significant similarity to the regions surrounding Kapi1-like prophages found in the NCBI database. The end of the first contig encodes the *dsdAXC* genes, while the end of the second contig encodes *yfdC*, *mlaA*, and *fadLIJ* genes. Primer pairs were designed to amplify the putative prophage-chromosome junctions (prophage and chromosome_left; prophage and chromosome_right primers, Table S2); PCR products were then sequenced and aligned with the original sequences to determine the integration site of Kapi1, and its orientation in the host chromosome. Kapi1 integrates into the chromosome between genes *yfdC* and *dsdC*, with phage *int* gene closest to the chromosomal *yfdC* locus, and the phage *xis* gene on the opposite end of the prophage closest to the

chromosomal *dsdC* gene (Figure 3A). Notably, the way in which the original MP1 sequence (11) was assembled actually captures the chromosome-prophage left_junction (Figure 3A), since the end of contig NZ_JEMI01000016 contains Kapi1 *xis* (annotated as *TorI* in the original sequence) directly downstream of *dsdC*. *yfdC* is a predicted inner membrane protein, belonging to the formate-nitrate transporter family and may be involved in resistance to surfactants (27); *dsdC* is a transcriptional regulator involved in D-serine detoxification (28). This region of the genome is hypervariable among different *E. coli* pathotypes, and a variety of prophage and phage-like genes are often found at this locus (29).

With the integration site for Kapi1 identified, we then looked back at our whole-genome sequencing data for KP7, the strain lacking the Kapi1 prophage. In this strain, the integration locus was correctly assembled; the two contigs that we confirmed to surround the integrated Kapi1 prophage in MP1 were assembled into one complete contig in KP7. We noticed a tRNA-Arg in between *dsdC* and *yfdC* that was not annotated on the contigs surrounding the Kapi1 prophage in MP1. Since tRNAs are common integration sites for phage (30), this site was further investigated. When we investigated the chromosomes of MP1 and MP13 with the Kapi1 prophage integrated as described above, we noticed a 17 bp duplication on either end of the integrated prophage; this sequence (5' – AATGGTGTCCCCTGCAG – 3') is found at the 3' end of the tRNA-Arg and is the putative Kapi1 *att* site (Figure 3A). To be clear, this tRNA-Arg is intact whether or not Kapi1 is integrated into the chromosome, since the 3' end is maintained by the Kapi1 putative *attP* when it integrates into the chromosome. The *attB* site was not picked up by the auto-annotation programs in MP1 and MP13 since during sequencing the two contigs surrounding Kapi1 were not assembled into the correct scaffold, as they were in KP7.

Interestingly, the Kapi1 putative *attP* is identical to prophages Sf6 (31), HK620 (26), and KplE1 (32) except for the 5' A which is excluded from the Sf6, HK620, and KplE1 *attP* sites. Like these phages, the Kapi1 *attP* lies between the *int* and *xis* genes, so upon integration into the host chromosome, the *int* and *xis* genes are located on either end of the prophage (Figure 3A). We generated a KP7 $\Delta attB::kan$ mutant resulting in a precise deletion of the 17 bp putative *attB*, and replacement with a kanamycin (kan) resistance cassette. Despite deletion of the putative *attB* site, Kapi1 retains the ability to infect and replicate within this host (Figure 3B). We noted a slight reduction in both the efficiency of plaquing and plaque size on this host compared to the WT (Figure 3B). We next determined the efficiency of lysogeny by Kapi1 of WT KP7 compared to KP7 $\Delta attB::kan$ by infecting each strain with Kapi1 at a multiplicity of infection of 10 and plating out survivor colonies. Using primers that span the prophage_left junction we consistently observed bands for WT KP7 (19/20 colonies screened), but never observed bands for KP7 $\Delta attB::kan$ (0/24 colonies screened). When these colonies were grown up overnight and their supernatant was spotted onto a lawn of naïve KP7, all KP7 cultures produced zones of clearing on KP7, and only 3/24 $\Delta attB::kan$ colonies produced zones of clearing. Since these $\Delta attB::kan$ survivor colonies did not appear to have any growth defect in comparison with WT KP7 survivor colonies, it seems unlikely that they were lytically infected with Kapi1, and we propose that rather they were lysogenized with Kapi1 via integration at an alternative *att* site elsewhere in the genome.

Although we did not explore this further experimentally, altered growth rate of the $\Delta attB::kan$ mutant, and/or increased latent period of Kapi1 when infecting this mutant could be responsible for the reduced plaque size (33). Similarly, the observed reduction in relative efficiency of plaquing could be explained by a reduced number of viable phage infections in the

mutant host, and/or any alterations to the latent period of Kapi1 that could result in it taking longer for plaques to become visible on the mutant host lawn compared to the WT (33). If integration is important for the lifecycle of Kapi1, it's possible that because Kapi1 cannot integrate into the mutant host chromosome at the same rate as the WT, this results in a longer latent period and thus smaller and fewer plaques on the mutant with the disrupted native *attB* site.

The Kapi1 receptor is LPS O-antigen, and is modified by Kapi1 both during initial binding and later during lysogeny

Many phage use LPS as a receptor for host cell infection, particularly *Podoviridae* (34), including phages HK620 (35) and P22 (24), among others. Based on this knowledge, we hypothesized that Kapi1 may use LPS as its receptor. We isolated spontaneous mutants resistant to Kapi1 by picking survivor colonies from lawns of KP7 overlaid with Kapi1. Since Kapi1 is temperate, survivor colonies were screened for lysogeny by colony PCR with primers spanning the prophage-chromosome junction, and by growing up survivor colonies overnight and spotting their supernatant onto KP7. Two Kapi1-resistant survivor colonies were verified to not be lysogenized by Kapi1, and were chosen for further analysis (KP61, KP62). LPS profiling by SDS-PAGE and silver staining showed that the Kapi1-resistant mutants have severely truncated LPS compared to the WT (Figure 4A). The genomes of KP61 and KP62 were sequenced, and variant calling was performed using WT KP7 as a reference. Both Kapi1-resistant strains contained only one mutation relative to WT; a 1 bp deletion in the *wzy* polymerase causing a frameshift resulting in a truncation (Figure 4B; Figure S3). Mutations in *wzy* result in synthesis of a complete core LPS, but only 1 O-unit is displayed on the cell surface (36), instead of the

usual long-chain O-antigens, consistent with what was observed on silver staining. These results indicate that Kapi1 likely binds the KP7 O-antigen as its primary receptor, however it appears that 1 O-unit is insufficient for recognition, as *wzy* mutants are completely resistant to infection (Figure 4C).

To confirm that O-antigen is the receptor for Kapi1, we proceeded to create two LPS mutants; KP7 $\Delta waaL::kan$ which has complete core, but no O-antigen (37), and KP7 $\Delta waaF::kan$ which has a severely truncated core structure consisting of lipid A, Kdo₂, and heptose (38). Surprisingly, KP7 $\Delta waaL::kan$ is able to be infected by Kapi1, but at an extremely low efficiency (faint clearing in undiluted spot, Figure 4C), whereas KP7 $\Delta waaF::kan$ is completely resistant to infection by Kapi1 (Figure 4C). This indicates that in the absence of O-antigen, Kapi1 may be able to recognize a portion of the outer-core structure that is intact in $\Delta waaL::kan$ but absent in $\Delta waaF::kan$, as a secondary receptor; in KP61 and KP62 the single O-unit may obscure this secondary receptor to prevent Kapi1 recognition/binding. Interestingly, *E. coli* W3110 *waaF* mutants have been shown to not produce flagella (39), so this presents another putative secondary receptor for Kapi1. However, no individual plaques can be observed when Kapi1 is spotted on KP7 $\Delta waaL::kan$ (Figure 4C) so it is also possible that the spot is the result of bactericidal activity, and not a productive phage infection (33). In light of the identification of the O-antigen as the receptor for Kapi1, our previous host-range results make sense; many of the strains we tested lack O-antigen (K-12 derivatives MC4100, MG1655, BW25113, W3110), and those that do produce O-antigen (natural isolates E2348/69, J96, DBS100, TJ-LM, TJ-WM, TJ-LR, Nissile 1917) do not appear to have the same O-antigen structure as KP7 on silver-stained SDS-PAGE (data not shown). Although the data presented here is not an extensive screen of all possible serotypes of *E. coli*, it is possible that Kapi1 is specific to one or a few serotypes;

identification of the precise region of O-antigen that Kapi1 recognizes as its receptor will help to clarify these results.

While performing LPS profiling in the previous experiments, we noticed that the LPS profiles of MP1 and MP13 differed from the LPS profile of KP7 (Figure 4A). We hypothesized that the change in LPS structure was due to lysogenic conversion by Kapi1. Several of the bacteriophage most similar to Kapi1 have been shown to cause seroconversion in their respective hosts. Phages Sf101 and Sf6 both encode O-acetyl transferases to cause seroconversion in their host, *S. flexneri* (25, 40); and phage P22 encodes an O-antigen glucosylation cassette (*gtrABC*) to cause seroconversion in its host *Salmonella enterica* serovar Typhimurium (41). We began by searching for CDS in the genome of Kapi1 with homology to known seroconversion proteins. Although none of the CDS in Kapi1 were annotated as possible O-antigen modifying/seroconverting proteins, closer examination revealed limited regions of homology to known seroconversion proteins in a few CDS. Kapi1 hypothetical protein 5 (*hyp5*) has limited homology with acyl- and acetyl- transferases, and hypothetical protein 19 (*hyp19*) has limited homology with an inhibitor of alpha polymerase (*iap*) involved in seroconversion. Many of the phages infecting *S. flexneri* that are closely related to Kapi1 encode seroconversion genes that are located near the *int-att-xis* region of the genome (9), so we included two additional CDS as putative seroconversion proteins due to their location in the genome and lack of homology to any known proteins in the NCBI viral BLASTp database; Kapi1 hypothetical protein 24 (*hyp24*) encoded between the tail spike protein and integrase, and hypothetical protein 25 (*hyp25*) encoded between the integrase and *attP* site.

We cloned each of these four putative seroconversion proteins into the pTrc99a overexpression vector (42), and introduced them into KP7, then compared the LPS profiles of

each overexpression strain, the vector control, and WT KP7 and MP13. Unfortunately, none of the strains overexpressing the four putative seroconversion proteins from Kapi1 had altered LPS profiles (data not shown). Although none of these proteins were found to be individually responsible for seroconversion, we could not eliminate the possibility that they could be working with other phage-encoded proteins to produce the altered LPS phenotype as there are numerous examples of seroconverting phage that encode entire gene cassettes responsible for seroconversion, including *Pseudomonas aeruginosa* phage D3 (43) and *Salmonella* phage P22 (41). We thus individually deleted *hyp5*, *hyp24*, and *hyp25* from Kapi1 lysogens and examined LPS profiles via silver stained SDS-PAGE; once again, none of these genes were found to be responsible for Kapi1-mediated O-antigen modification, as the LPS profiles of each mutant were identical to the WT lysogen (data not shown). Notably, we were unable to delete *hyp19*. Both Top10 (Invitrogen) cells carrying pUC18 (44) with the $\Delta hyp19$ construct, and our donor strain MFDpir (45) carrying pRE112 (46) with the $\Delta hyp19$ construct had quite severe growth defects. We hypothesize that this may be due to one of the hypothetical proteins flanking *hyp19* that is present on the deletion construct exerting some type of toxicity in these strain backgrounds which are not lysogenic for Kapi1. Additionally, when screening sucrose-resistant colonies for double-crossovers, we noted an unusually high rate of Kapi1 excision. Colonies with single-crossovers (verified to be lysogenic via colony PCR) were grown in Luria-Bertani broth (LB) for 6 hours before plating on sucrose to select against pRE112. Only 9/2000 sucrose-resistant colonies retained the Kapi1 prophage (0.45%), while under normal conditions, approximately 90% of the population retains the Kapi1 prophage after 24 hours of growth in LB (data not shown).

Although we were unsuccessful in identifying the gene responsible for Kapi1-mediated O-antigen modification, it is interesting that the genetic basis for this O-antigen modification by Kapi1 appears to not be well-conserved, as no strong hits to known O-antigen modifying proteins could be identified. It will be valuable to determine the molecular mechanisms behind Kapi1 phage-mediated O-antigen modification, and whether these mechanisms are indeed novel. Since we identified the LPS O-antigen as the receptor for Kapi1, and Kapi1 lysogens are protected from superinfection by Kapi1, there is a distinct possibility that Kapi1-mediated O-antigen modification is a mechanism for the observed superinfection immunity. If the modified LPS displayed on the cell surface of Kapi1 lysogens is unable to be bound by Kapi1 phage, this could prevent superinfection by Kapi1 or similar O-antigen binding phages. Besides superinfection immunity, modification of the LPS O-antigen could alter the immunogenicity of Kapi1 lysogens, if for example this modification leads to a change in serotype. However, what the relevance of this would be for a non-pathogenic strain of *E. coli* such as MP1 is unclear. Finally, modification of the O-antigen structure could alter the barrier function of LPS, resulting in, for example altered resistance to antibiotics or altered membrane rigidity (10).

Phages that use the O-antigen as a receptor also commonly modify the O-antigen using the tail spike protein (*tsp*) to facilitate movement to the bacterial outer membrane, where irreversible binding and particle opening can occur (47). To determine if O-antigen degradation via *tsp* was responsible for the altered LPS structure between Kapi1 lysogens and non-lysogens observed on SDS-PAGE (Figure 4A), we set up a mock infection with Kapi1 and purified KP7 LPS and ran these samples alongside the uninfected controls from KP7 and MP13. It appears that treatment of purified KP7 LPS with Kapi1 results in an altered LPS structure (suggesting degradation of O-antigen by Kapi1), but this structure is different from that of either the WT or

lysogenic backgrounds (Figure 4D). Therefore, Kapi1 is responsible for alteration of the LPS structure both upon binding O-antigen prior to infection, and later via lysogenic conversion. Kapi1 *tsp* has considerable sequence conservation in the head-binding domain (most similar to phages HK620 and Sf101), but no sequence similarity to any other viral proteins in the NCBI database along the length of the protein. Therefore, it is difficult to predict what type of enzymatic activity the Kapi1 *tsp* may have, and further work is needed to characterize the molecular mechanisms of Kapi1-mediated O-antigen modification through *tsp*.

Kapi1 is an unstable temperate phage, and shows an altered lifestyle in simulated intestinal conditions

To investigate the lifestyle of Kapi1, we measured the titer of Kapi1 along with host cell counts in cultures of lysogens. After only 24 hours of growth, $\sim 1 \times 10^8$ PFU/mL can be isolated from standard laboratory cultures of lysogens (data not shown). The ratio of phage per cell (PFU/CFU) in cultures of lysogens is only 0.050 ± 0.020 after 24 hours of growth and rises to 41.66 ± 11.79 PFU/CFU when identical cultures are grown with sub-inhibitory concentrations of mitomycin C (Figure 5A). This suggests that Kapi1 can be induced through the traditional SOS pathway in response to DNA damaging agents, although there appears to be a basal level of spontaneous induction even in the absence of mitomycin C. To determine if Kapi1 lytic replication is *recA*-dependent, we repeated the above assay in a $\Delta recA::kan$ lysogen background and did not observe plaque formation in either the presence or absence of mitomycin C (data not shown), indicating that both spontaneous and DNA-damage-response induction of Kapi1 occur in a *recA*-dependent manner. Given the level of spontaneous induction, we wondered how

carriage of Kapi1 might impact host cell growth rate. Kapi1 slightly reduces the growth rate of its host compared to the identical non-lysogenized strain, KP7 (Figure S4A).

Since MP1 was recently isolated from the feces of a healthy mouse (11) and is more host-adapted than our standard laboratory strains of *E. coli* such as MG1655 or MC4100, we wondered if Kapi1 might be important to the biology of commensal *E. coli* in the GI tract. To investigate the biology of Kapi1 under more physiologically-relevant conditions, we repeated the same experiments following the numbers of phage and host cells in media composed of 50% LB and 50% simulated intestinal fluid (SIF) as well as 50% LB and 50% simulated gastric fluid (SGF) (48). We found lower ratios of Kapi1 PFU to host CFU in SIF when compared to LB (Figure 5A), although the strains grow to nearly identical cell densities. Importantly, the rate of Kapi1 adsorption to host cells is equal across all tested media types (Figure S4B); thus, we conclude that the differences in PFU/CFU ratios reflect the approximate induction rates of Kapi1. We next performed prophage stability assays, which showed that upon repeated sub-culturing of a lysogen, the percentage of the population carrying Kapi1 reduces by approximately 10% upon each successive sub-culture in LB, and that the lysogen population is more stable in media composed of 50% LB and 50% SIF (Figure 5B). This result supports our previous finding that the original MP7 isolate lost the Kapi1 prophage before arriving at our lab, likely during handling or passaging. Finally, we monitored the expression of Kapi1 *CI* using a luminescent reporter assay. The regulatory region of the Kapi1 genome is well-conserved, and is reminiscent of phage λ , which famously uses the *CI* phage repressor to maintain lysogeny (reviewed by (49)). We thus cloned the promoter region of Kapi1 *CI* (-86 to +13) into the luminescent reporter plasmid pNLP10 (50), and monitored light production as an indicator of *CI* expression, and thus maintenance of the lysogenic cycle. Monitoring *CI-lux* activity over time revealed interesting

gene expression patterns; we found that when Kapi1 lysogens are grown in SIF, a clear “decision” is made around 2 hours post-subculture to favor lysogeny, as seen by a sharp increase in *CI-lux* activity (Figure 5C). In contrast, the *CI-lux* expression patterns in LB and SGF do not show this induction and instead plateau throughout exponential growth (Figure 5C). We also included a culture grown with sub-inhibitory concentrations of mitomycin C, and as expected, noted the lowest *CI-lux* activity in this condition (Figure 5C). After 24 hours of growth, we again measured *CI-lux* activity and observed increased *CI-lux* in SIF compared to both LB and SGF (Figure 5D), supporting our previous results (Figure 5ABC).

The previous three assays agree; there is a lower level of Kapi1 induction when lysogens are grown in LB supplemented with SIF compared to LB alone or LB supplemented with SGF, and this low level of induction is likely responsible for the increased stability of the lysogen population in media supplemented with SIF. This result is unexpected because LB is considered a non-stressful standard lab media, yet there is a higher proportion of spontaneous phage induction in this condition, compared to a less rich and more challenging simulated intestinal media (Figure S4C). Additionally, since the PFU/CFU in LB diluted 50% with either distilled water or phosphate-buffered saline (PBS) were nearly identical to full LB (Figure S4D), it seems likely that the lower rate of induction observed in SIF is not simply due to dilution of the rich LB media and may be specific to intestinal conditions. Since the rate of Kapi1 induction is lower in SIF, and more lysogens are retained in the population in SIF (Figure 5), this could indicate that lysogeny with Kapi1 is selected for in intestinal conditions, as it is retained in the prophage state in SIF. This could mean that when integrated as a prophage, Kapi1 may provide some advantage to the cell which may be dispensable under standard lab conditions. Prophages have been shown to provide fitness benefits to their hosts, including resistance to osmotic,

oxidative and acid stresses, as well as influencing biofilm formation (51); many of these stresses would be encountered during colonization of the mammalian GI tract. Notably, there are also examples where prophages are detrimental to their hosts' overall fitness during colonization of the GI tract (52); our observations regarding the stability of Kapi1 lysogens in simulated intestinal conditions warranted further investigation into how (or if) Kapi1 influences the fitness of commensal *E. coli* during colonization of the GI tract.

Kapi1 confers a competitive advantage to its host

To investigate how carriage of Kapi1 might impact the overall fitness of its host, we performed competitive assays between KP7 and MP13. *In vitro*, we co-cultured KP7 and MP13 in either LB or SIF, plating cultures onto tetracycline and enumerating the number of red-fluorescent (KP7), and green-fluorescent (MP13) colonies every 24 hours. MP13 strongly out-competes KP7, and this competitive advantage is further enhanced when cells are co-cultured in SIF, compared to LB (Figure 6A). Although the difference in the relative proportion of each strain between LB and SIF after 48 hours of co-culture is not statistically significant, these results were replicable (data not shown). We expected that MP13 may out-compete KP7 since it carries a phage that KP7 is sensitive to, but were surprised by the strength of this competition, since Kapi1 is a temperate phage. Further, the increased strength of Kapi1 lysogen competitiveness in SIF supports our previous findings regarding the stability of Kapi1 lysogens in SIF (Figure 5), suggesting that the interactions between Kapi1 and its host are altered in simulated intestinal conditions, compared to standard lab conditions. These findings open the doors to ask meaningful questions about what the precise roles of temperate phages are in the GI

tract, and how their interactions with their bacterial hosts are altered in this environment, compared to a laboratory environment.

To expand on these findings, we performed competitive assays *in vivo*, utilizing specific-pathogen-free (SPF) mice to study these interactions in the presence of a healthy microbiome. We performed two treatments: mice were treated with either KP7 or MP13 alone (colonization) or a mixture of KP7 and MP13 (competitive colonization). Both KP7 and MP13 stably colonize SPF mice for at least 28 days without significantly impacting mouse body weight (data not shown), but there is no difference in colonization levels between the two strains (Figure 6B). Interestingly, this data is contradictory to what others have found regarding the fitness of lysogens in the GI tract (52, 53). Importantly, both of these previous studies were performed in germ-free mice, while we studied the fitness of lysogens in the presence of a complex microbiome. Further, it is not surprising that each unique phage-host pair may interact with each other and with their surrounding environment in unique ways. These findings highlight the knowledge gap in our current understanding of how temperate phage behave in physiologically-relevant conditions. During competitive colonization of SPF mice, MP13 once again strongly out-competes KP7 (Figure 6C), supporting our *in vitro* findings. In contrast to our colonization results (Figure 6B) the competitive colonization results (Figure 6C) are supported by the previously mentioned studies (52, 53) which found that temperate phage provided a competitive advantage to their host when in competition with naïve phage-susceptible hosts. Our findings that the role of Kapi1 in colonization appears to be relevant only when in competition with non-lysogenized neighbors is a novel finding in the realm of *in vivo* studies investigating the role of temperate phages. Importantly, competition is a physiologically-relevant condition, as bacteria are never “alone” when colonizing the healthy GI tract; these results warrant further

investigation into how temperate bacteriophage modulate community dynamics in these densely colonized environments. Our previous results (Figure 5) demonstrate that in pure culture, lysogeny (as opposed to lytic induction of Kapi1) is favored in SIF, but whether or not this is maintained during competition is still under investigation. It is tempting to assume that Kapi1 provides a competitive advantage to its host via direct killing of naïve competitors, but if lysogeny is favored in the intestine as our previous results suggest, then this may not be the mechanism of competition.

In conclusion, we have isolated and begun the characterization of a novel bacteriophage infecting commensal *E. coli*. The genome of Kapi1 has been sequenced and annotated, and the prophage integration site in the host genome has been identified. Further, we demonstrate that Kapi1 shows unstable lysogeny, and that lysogeny appears to be selected for in intestinal conditions. O-antigen is the Kapi1 receptor and Kapi1 appears to modify the host O-antigen upon initial binding, and later in infection through lysogenic conversion, although the molecular mechanisms are yet to be elucidated. Our findings show that Kapi1 lysogeny confers a competitive advantage during colonization of the mouse GI tract, and we propose that MP1 and Kapi1 will serve as a good model system to explore what role(s) temperate phages may play in colonization of the GI tract by commensal strains of *E. coli*.

Materials and Methods

Bacterial Strains and Growth Conditions

Strains MP1, MP7, and MP13 were a generous gift from the Goulian lab (11). A complete list of bacterial strains used in this study can be found in Table S1. Unless otherwise

specified, all strains were grown in LB (10 g/L tryptone, 5 g/L NaCl, 5 g/L yeast extract) supplemented with the appropriate antibiotics, at 37 °C and 225 rpm. When plated, cells were grown on LB 1.5% agar supplemented with the appropriate antibiotics and incubated inverted at 37 °C. Antibiotic concentrations used are as follows: ampicillin 100 µg/mL, kanamycin 30 or 50 µg/mL, tetracycline 15 µg/mL. All chemicals were obtained from MilliporeSigma Canada.

Phage Isolation, Propagation, Host Range, and Transmission Electron Microscopy

Phage were isolated from overnight cultures of MP1 or MP13 by pelleting cells, and filter-sterilizing the supernatant using a 0.45 µm syringe-driven filter. Individual plaques were isolated, propagated, and phage stocks prepared by previously described methods (55), slightly modified. Briefly, the above phage-containing supernatant was mixed with susceptible host strain KP7 1:1, then 3.5 mL soft agar (LB 0.7% agar) was added, the mixture was poured onto solid LB agar plates, and incubated overnight. Individual plaques were picked with a sterile Pasteur pipette and gently resuspended in 500 µL modified suspension media (SM) (50 mM Tris-HCl (pH 7.4), 100 mM NaCl, 10 mM MgSO₄). This suspension was then amplified using the soft-agar overlay technique as above. Plates with near-confluent lysis were used to prepare high-titre stocks by collecting the soft-agar layer as follows: SM was poured onto the surface of the plate, and soft agar was gently scraped into a 50 mL Falcon tube using a sterile scoopula, rocked at room temperature for 1 hour and centrifuged to pellet the soft agar. The supernatant was filter-sterilized using a 0.22 µm syringe-driven filter, and stored at 4 °C. Plaque diameter was measured from 10 plaques, using Fiji software (56); the mean and standard deviation are reported.

Phage samples were prepared for transmission electron microscopy and imaged by previously described methods using uranyl acetate as a background stain (55) at the University of Alberta Advanced Microscopy Facility. Virion measurements were performed using Fiji software (56) from 44 phage particles; the mean and standard deviation are reported.

The host range for Kapi1 was determined by growing strains of interest overnight, then 50 μ L of overnight culture was added to 3.5 mL soft agar and poured onto a LB plate. Once solidified, Kapi1 lysate was serially diluted and spotted onto each strain. In parallel, a whole plate overlay was prepared as above, using 300 μ L undiluted phage lysate and 50 μ L overnight culture. The following day, plates were scored for presence or absence of plaques, with KP7 included as a positive control. The whole plate overlays were collected as above, for each strain tested. These “trained” Kapi1 lysates were serially diluted and spotted back onto the same strain to see if Kapi1 host range could be expanded by extended incubation with a particular host, as compared to the first round of spotting.

Genome Sequencing and PCR

Pure cultures were grown by streaking from glycerol cryo-stocks, then picking single colonies and growing overnight. DNA was extracted from overnight cultures using the Lucigen MasterPure Complete DNA and RNA Purification Kit, and the concentration and quality of gDNA was checked using the NanoDrop 2000c. Library preparation and whole genome sequencing was performed by the Microbial Genome Sequencing Centre (MiGS, Pittsburgh PA). Libraries were prepared with the Illumina Nextera kit, and sequenced using the NextSeq 550 platform.

PCR was performed using Taq polymerase (Invitrogen) following the manufacturer's directions; a single colony was suspended in 20 μ L nuclease-free water and 5 μ L of this suspension was used per 50 μ L reaction. Phage DNA was extracted via phenol/chloroform extraction (57), and 1-5 μ L of phage DNA was used per 50 μ L reaction. PCR products were checked by running 10 μ L on a 2% agarose gel and staining with ethidium bromide. All primers used in this study can be found in Table S2. Sanger sequencing was performed by the Molecular Biology Service Unit at the University of Alberta.

Genome Assembly, Annotation, and Taxonomy

Paired-end reads of 2x150 bp received from MiGS were uploaded to the public server at usegalaxy.org (58). Galaxy, and all tools there-in, was used for bacterial genome assembly, annotation, and analysis as follows. Illumina adapters and low-quality reads were trimmed using Trim Galore! (59). Reads were then checked for quality using FastQC (60) and MultiQC (61). Trimmed reads were *de novo* assembled using Unicycler (62), functioning as a SPAdes (63) optimizer as no long-read data was generated. Quality of assemblies was assessed using Quast (64), and bacterial genomes were then annotated using Prokka (65, 66). SnapGene software (from Insightful Science; available at snapgene.com) was used for genome visualization, creating genome maps, and designing primers.

The Kapi1 genome was manually annotated using previously described methods (67), slightly modified. This method uses a rigorous approach to score CDS based on various parameters; low-scoring CDS are then discarded, and the remaining CDS are analyzed to determine their correct start codon, again based on a scoring system. Briefly, the prophage

genome was run through three auto-annotation programs; GenemarkS (68), Glimmer3 (69) at CPT Phage Galaxy public server (cpt.tamu.edu/galaxy-pub), and Prokka (65) at usegalaxy.org (66). The coding potential for each putative CDS was determined using GenemarkS coding potential graph (68). Putative CDS were then searched against NCBI's non-redundant protein database using BLASTp (20), and scored based on whether they had significantly similar hits (as determined by the E-value, using a cut-off of E-10), and whether those hits were known proteins or hypotheticals. Each CDS was scored based on the length of overlap with neighboring CDS, as extremely long overlaps are unlikely, while short overlaps of 1, 4 or 8 bp are more favorable as these suggest organization into an operon. Finally, CDS were scored based on the length of the ORF, where extremely short CDS are penalized the most heavily. Low scoring CDS were discarded prior to start codon identification. Start codons for each CDS were scored in a similar manner, using again the coding potential graph from GenemarkS (68), the number of auto-annotation programs that selected the start codon, sequence similarity matches in NCBI, and length of the ORF. These parameters are listed in order of most important to least and were used to select the most likely start codon for each CDS. CD-Search (70) was also performed for all CDS to assist with functional assignment.

Taxonomic evaluation was performed using vContact2 v9.8 (23) through the CyVerse platform (www.cyverse.org). The analysis was run with the default parameters, using NCBI Bacterial and Archaeal Viral RefSeq V85 (with ICTV + NCBI taxonomy) as the reference database. The resulting network was visualized using CytoScape (71). Duplicated edges were removed from the network (edges represent connections between two nodes, in this case viruses), and only first-neighbors to Kapi1 were kept (nodes that have a direct connection to the Kapi1 node). An edge-weighted, spring-embedded layout was used so that nodes that are more

closely related appear closer together spatially, and edges were weighted so that stronger connections (ie. more sequence similarity between two viruses) appear darker and thicker.

Lipopolysaccharide Profiling

The general structure of LPS was analyzed by profiling LPS extracts on SDS-PAGE with silver staining. LPS was extracted using a modified proteinase K micro-digestion protocol (72) as follows; bacterial strains of interest were grown overnight, then 1 mL of this culture was washed twice with PBS, resuspended to a final OD⁶⁰⁰ of 2.0 in PBS, and pelleted. Cells were resuspended in 50 µL lysis buffer (2% SDS, 4% 2-mercaptoethanol, 10% glycerol, 1M Tris pH 6.8, bromophenol blue to a deep blue color) and incubated at 95 °C for 10 minutes to lyse cells. Whole cell lysate was cooled to room temperature, then 10 µL 2.5 mg/mL proteinase K (20 mg/mL stock solution was diluted in lysis buffer first) was added and incubated at 56 °C for 1 hour. Standard polyacrylamide gels were prepared with 12% acrylamide (19:1 acrylamide:bisacrylamide) (73), 1-5 µL of proteinase K-treated whole cell lysate (LPS extract) was loaded per well and run at 80 V in Tris-glycine running buffer (25 mM Tris, 200 mM glycine, 0.1% SDS) until the dye front nearly reached the bottom of the gel. Silver staining was performed as previously described (74), and imaged on a clear petri dish using an iPhone camera.

To determine if Kapi1 is capable of degrading LPS, LPS was extracted from cultures of KP7 (non-lysogenic for Kapi1) and MP13 Kapi1 lysogens using a phenol extraction method adapted from Davis and Goldberg (75), purified KP7 LPS was then incubated with Kapi1 and compared to the untreated controls. The phenol extraction was used in place of proteinase K

digestion (as above), as Kapi1 was not viable in the lysis buffer, even after inactivation of proteinase K at high temperatures (data not shown). Briefly, KP7 and MP13 were grown up overnight, cultures were pelleted and washed in PBS, then resuspended in 1.5 mL PBS to give a final OD⁶⁰⁰ of 2.0. Cells were pelleted and resuspended in 200 µL Laemmli buffer (50 mM Tris-HCl pH 6.8, 4% SDS, 10% glycerol, 0.1% bromophenol blue, 5% β-mercaptoethanol), then boiled for 15 minutes to lyse cells. Once cool, DNase I and RNase A were added to cell lysates at 37 °C for 10 minutes. Proteinase K was then added, and lysates were incubated at 55 °C overnight. The following day, 200 µL Tris-saturated phenol was added to lysates, then vortexed for 10 seconds before incubating at 65 °C for 15 minutes. Once cool, samples were centrifuged at 14,000 rpm for 10 min at 4 °C, and the upper phase was transferred to a new tube. 2.5 vol ethanol was added to precipitate LPS, then centrifuged at 15,000 rpm for 20 minutes. Supernatant was discarded, pellets air dried, then resuspended in 50 µL nuclease-free water. Kapi1 was added to KP7 LPS at a “MOI” of 10 (assuming that 1.5 mL of OD⁶⁰⁰ 2.0 culture was concentrated into 50 µL) and incubated at 37 °C without shaking for 30 minutes. Phage-treated and untreated KP7 LPS, along with MP13 untreated LPS were run on SDS-PAGE and silver stained, as above, using 15 µL of LPS extract, as this extraction method produced a lower yield. A portion of the phage-treated samples were serially diluted and spotted onto an KP7 soft-agar overlay to ensure that phage particles remained viable after incubation with LPS (data not shown).

Generation of Mutants

Spontaneous Kapi1-resistant mutants were isolated by spotting Kapi1 onto a soft-agar overlay prepared with KP7. The following day, six colonies that grew within the cleared phage

spots were picked and re-streaked onto LB. The following day, these colonies were screened for lysogeny with Kapi1 using PCR with primers that span the phage-genome junction. Two colonies that did not produce phage-genome bands were grown up overnight, then the supernatant was filter-sterilized and spotted onto a soft-agar overlay prepared with KP7. The absence of plaques on KP7 confirmed that these two mutants are not Kapi1 lysogens. To identify which mutations were responsible for the Kapi1-resistant phenotype, each colony was sent for whole genome sequencing at MiGS, as above. SNPs were identified using BreSeq (76) and snippy (77) through Galaxy (58), using the KP7 genome as a reference.

KP7 $\Delta waaF::kan$ and MP13 $\Delta recA::kan$ mutants were constructed via P1 transduction, as previously described (78, 79) using the corresponding Keio collection mutant (80) as a donor. The previously described Lambda Red system (81) was used to generate KP7 $\Delta waaL::kan$. Once sequencing confirmed the correct mutation in *waaL*, P1 transduction was used (as above) to move these mutations into a fresh KP7 background, to avoid the possibility of any off-site mutations acquired during construction. Putative seroconversion genes were deleted via allelic exchange, as previously described (82).

Efficiency of lysogeny, phage and cell counts, prophage stability, and luminescent reporter assays

Efficiency of lysogeny was determined by infecting KP7 and KP7 $\Delta attB::kan$ with Kapi1 at a multiplicity of infection of 10 at 37 °C for 30 minutes, then washing and plating out surviving cells. The following day, survivor colonies were screened by colony PCR using primers that span the prophage_left junction (chromosome_left, prophage_left primers, Table

S2) to look for integration of Kapi1 at the putative *attB* site. The same set of colonies were also grown overnight in liquid culture. Overnight cultures were then assayed for lysogeny by spotting the culture supernatant onto a soft-agar overlay prepared with KP7; the supernatants of Kapi1 lysogens produce plaques on KP7.

To monitor the CFUs and PFUs in cultures of MP13 Kapi1 lysogens over time, three colonies of MP13 and three colonies of KP7 (non-lysogenic control) were picked and grown overnight. The next day cultures were adjusted to OD⁶⁰⁰ 1.0 to ensure equal cell numbers, then sub-cultured 1:100 into LB, LB with 0.5 ng/μL mitomycin C, LB mixed 50-50 with simulated intestinal fluid (SIF – 6.8 g KH₂PO₄, 1.25 g pancreatin, 3 g bile salts in 1 L dH₂O, pH adjusted to 7 (48)), or LB mixed 50-50 with simulated gastric fluid (SGF – 2 g/L NaCl, 3.2 g/L porcine mucosa pepsin, pH adjusted to 3.5 (48)). After 24 hours incubation, an aliquot was taken from each culture. Cells were spun down, washed in PBS (137 mM NaCl, 2.7 mM KCl, 10 mM Na₂HPO₄, 1.8 mM KH₂PO₄), serially diluted, plated onto LB, and grown overnight to enumerate the number of cells in the culture. In parallel, the culture supernatants (containing phage) were serially diluted, spotted onto KP121 soft-agar overlays, and incubated overnight at 30 °C to enumerate the number of phage particles in the culture. KP121 was used to enumerate phage as this strain has a much lower efficiency of lysogeny with Kapi1 compared to WT KP7, allowing for more accurate enumeration.

Prophage stability was assayed by serially propagating cultures of Kapi1 lysogens. MP13 cultures were grown in biological triplicates (three independent colonies) in either LB, LB mixed 50-50 with SGF, LB mixed 50-50 with SIF, or LB with 0.5 ng/μL mitomycin C for 24 hours, then sub-cultured into fresh media and grown for another 24 hours. After a total of 48 hours incubation (1 passage) cells were spun down and washed twice in PBS, serially diluted, plated on

LB and grown overnight to get individual colonies. Using velvet squares, colonies were replicated onto a LB plate spread with 50 μ L of an overnight culture of KP7 to screen for lysogens (83). This replica plating technique results in two distinct phenotypes: colonies that produce a zone of clearing in the KP7 lawn are scored as lysogens, and colonies without a zone of clearing are scored as non-lysogenic (see Figure S5 for examples and experimental verification of these phenotypes). The relative loss of Kapi1 lysogens in the culture was calculated by dividing the non-lysogen CFU/mL by the total CFU/mL.

Luminescent reporters were constructed by cloning the *CI* promoter region into the luminescent reporter plasmid pNLP10 (50) via standard PCR cloning. Kapi1 lysogens carrying pNLP10 *CI-lux* were grown overnight in biological triplicate in LB + Kan. The next day, each overnight culture was sub-cultured 1:50 into LB + Kan in 4 replicates (one for each media type), so that each media type assayed contained the same 3 biological replicates. After 1.5 hr growth, cultures were spun down for 5 minutes at 4000 rpm. Supernatant was poured off, and cells were resuspended in 2 mL of either: LB, 50% LB / 50% SGF, 50% LB / 50% SIF, or LB + 0.5 ng/ μ L mitomycin C. To measure luminescence, 100 μ L of culture was aliquoted into a sterile black 96-well plate, then luminescence and absorbance at 600 nm were measured using the Victor X3 2020 multilabel plate reader (Perkin Elmer). Luminescence (counts per second, CPS) was divided by bacterial optical density (OD^{600}) to account for variations in growth, and the average of 3 biological replicates was plotted, with the standard deviation shown as error bars. The experiment was repeated 3 times.

Bacterial competitions

In vitro competitive assays were performed by growing KP7 and MP13 overnight cultures in biological triplicate. Each culture was standardized to an OD⁶⁰⁰ of 1.0 in PBS to ensure approximately equal cell numbers, then 10 µL of KP7 and 10 µL MP13 were sub-cultured into 2 mL of either LB or SIF. Immediately, co-cultures were briefly vortexed and 100 µL of the culture was aliquoted, then co-cultures were incubated at 37 °C 225 rpm for 24 hours. To enumerate the starting cell counts of each strain, samples from co-cultures were serially diluted 10-fold in PBS, then 10 µL of each dilution was plated onto LB containing 15 µg/mL tetracycline using the track dilution method; once dry, plates were incubated overnight at 30 °C. Plates were imaged using the ChemiDoc system, using DyLight 550 for visualization of *mcherry*, and StarBright B520 for visualization of *gfpmut3.1*. Notably, although cultures were standardized to equal optical density, the starting inoculum of KP7 was much higher than MP13. We hypothesize that this is due to the demonstrated spontaneous induction of Kapi1; the resulting cell lysis in a sub-set of the MP13 population could result in more cellular debris and thus a lower number of viable cells at the same optical density measurement. Every 24 hours, co-cultures were sub-cultured 1:100 into fresh media, and again 100 µL of the mature culture was sampled to perform cell counts as above.

In vivo studies were performed using mature adult C57BL/6J male mice kept in the animal facility at the University of Alberta. Mice not harboring coliforms, as confirmed by plating on MacConkey agar (BD, Sparks, MD), were housed on aspen wood chip bedding materials in sterilized filter-topped IsoCages with nestlets, mouse huts and nesting materials as enhancements. The room environment was controlled for temperature (20 - 22 °C), relative humidity (40 %), and light cycle (12-h light and 12-h darkness). Mice were given *ad libitum* access to water and a standard chow diet and cages were handled in a biosafety cabinet under

specific pathogen-free conditions. Animals were randomly grouped into 4 to 5 mice per cage by a blinded lab animal technician. Cages were allocated into 3 treatments: colonization with KP7, colonization with MP13, and competitive colonization with a mixture of KP7 and MP13. KP7 and MP13 were cultivated in 10 mL of LB medium (Fisher Scientific, Nepean, ON) at 37 °C for 16 h. Each mouse received 0.1 mL of culture medium containing approximately 1.0×10^7 CFU of *E. coli* cells by oral gavage. Body weights were recorded, and fecal samples were collected at 0, 7, 14, and 28 days post-infection (DPI). Enumeration of *E. coli* was conducted by serial dilutions of fecal samples plated on MacConkey agar and total CFUs per gram of feces were then calculated. For competitive colonization, MacConkey plates were replica-plated onto LB + Tet and imaged as above. The protocols employed were approved by the University of Alberta's Animal Care Committee and in accordance with the guideline of the Canadian Council on the Use of Laboratory Animals.

Statistics and Data Visualization

The Shapiro-Wilk test was used to check the normality of data distribution and the log transformation or arcsin transformation was applied to address skewed data. Student's t-test or one-way ANOVA with Tukey's post-hoc test was performed accordingly. Data were present as mean \pm standard deviation (SD). A p-value below 0.05 was considered statistically significant. Unless otherwise specified, statistical analysis was performed using either Microsoft Excel or GraphPad Prism 7 (54), and graphs were generated in Microsoft Excel.

Data Availability

The genome of Kapi1 can be accessed from NCBI GenBank (accession MT813197).

Acknowledgments

The authors thank Arlene Oatway from the University of Alberta Advanced Microscopy Facility for assistance with transmission electron microscopy; the University of Alberta Molecular Biology Facility for assistance with Sanger sequencing; the Microbial Genome Sequencing Centre for assistance with whole genome sequencing; Dr. Mark Goulian for critical reading of the manuscript and for strains MP1, MP7, MP13; Jaclyn McCutcheon for advice on phage isolation and preliminary characterization; Dr. Brent Weber for advice on identification of the integration site; Ashley Gilliland for thoughtful discussion of results and data analysis; and Stephanie Tollenaar for assistance with animal work.

This research was supported by operating grants from The National Sciences and Engineering Research Council (NSERC) and The Canadian Institutes of Health Research (CIHR), and a project grant from the AMR – One Health Consortium, funded by the Major Innovation Fund program of the Ministry of Jobs, Economy and Innovation, Government of Alberta, to TR. BPW is supported by the Canada Research Chairs program. KP was supported by an NSERC Alexander Graham Bell Canada Graduate Scholarship - Master's, Walter H Johns Graduate Fellowship, University of Alberta Science Graduate Scholarship, and Susan Eberlein Graduate Scholarship in Genetics.

References

1. Tenaillon O, Skurnik D, Picard B, Denamur E. 2010. The population genetics of commensal *Escherichia coli*. *Nat Rev Microbiol* 8:207–217.
2. Howard-Varona C, Hargreaves KR, Abedon ST, Sullivan MB. 2017. Lysogeny in nature: Mechanisms, impact and ecology of temperate phages. *ISME J* 11:1511–1520.
3. Touchon M, Bernheim A, Rocha EPC. 2016. Genetic and life-history traits associated with the distribution of prophages in bacteria. *ISME J* 10:2744–2754.
4. Kim MS, Bae JW. 2018. Lysogeny is prevalent and widely distributed in the murine gut microbiota. *ISME J* 12:1127–1141.
5. Hsu BB, Gibson TE, Yeliseyev V, Liu Q, Lyon L, Bry L, Silver PA, Gerber GK. 2019. Dynamic Modulation of the Gut Microbiota and Metabolome by Bacteriophages in a Mouse Model. *Cell Host Microbe* 25:803-814.e5.
6. Frazão N, Sousa A, Lässig M, Gordo I. 2019. Horizontal gene transfer overrides mutation in *Escherichia coli* colonizing the mammalian gut. *Proc Natl Acad Sci U S A* 116:17906–17915.
7. Davies E V., Winstanley C, Fothergill JL, James CE. 2016. The role of temperate bacteriophages in bacterial infection. *FEMS Microbiol Lett*.
8. Van Bellegheem JD, Dąbrowska K, Vaneechoutte M, Barr JJ, Bollyky PL. 2019. Interactions between bacteriophage, bacteria, and the mammalian immune system. *Viruses* 11.
9. Allison GE, Verma NK. 2000. Serotype-converting bacteriophages and O-antigen

modification in *Shigella flexneri*. Trends Microbiol 8:17–22.

10. Guest RL, Rutherford ST, Silhavy TJ. 2020. Border Control : Regulating LPS Biogenesis. Trends Microbiol 1–12.

11. Lasaro M, Liu Z, Bishar R, Kelly K, Chattopadhyay S, Paul S, Sokurenko E, Zhu J, Goulian M. 2014. Escherichia coli isolate for studying colonization of the mouse intestine and its application to two-component signaling knockouts. J Bacteriol 196:1723–1732.

12. Arndt D, Grant JR, Marcu A, Sajed T, Pon A, Liang Y, Wishart DS. 2016. PHASTER: a better, faster version of the PHAST phage search tool. Nucleic Acids Res 44:W16–W21.

13. Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS. 2011. PHAST: A Fast Phage Search Tool. Nucleic Acids Res 39:347–352.

14. Schoch CL, Ciufo S, Domrachev M, Hottel CL, Kannan S, Khovanskaya R, Leipe D, McVeigh R, O'Neill K, Robertse B, Sharma S, Soussov V, Sullivan JP, Sun L, Turner S, Karsch-Mizrachi I. 2020. NCBI Taxonomy: A comprehensive update on curation, resources and tools. Database 2020:1–21.

15. Ju T, Shoblak Y, Gao Y, Yang K, Fohse J, Finlay B, Wing So Y, Stothard P, Willing BP. 2017. Initial Gut Microbial Composition as a Key Factor Driving Host Response to Antibiotic Treatment, as Exemplified by the Presence or Absence of Commensal Escherichia coli. Appl Environ Microbiol 83.

16. Hull RA, Gill RE, Hsu P, Minshew BH, Falkow S. 1981. Construction and expression of recombinant plasmids encoding type 1 or D-mannose-resistant pili from a urinary tract infection Escherichia coli isolate. Infect Immun 33:933–938.

- 826 17. Iguchi A, Thomson NR, Ogura Y, Saunders D, Ooka T, Henderson IR, Harris D,
827 Asadulghani M, Kurokawa K, Dean P, Kenny B, Quail MA, Thurston S, Dougan G,
828 Hayashi T, Parkhill J, Frankel G. 2009. Complete genome sequence and comparative
829 genome analysis of enteropathogenic Escherichia coli O127:H6 strain E2348/69. J
830 Bacteriol 91:347–354.
- 831 18. Nissle A. 1918. Die antagonistische Behandlung chronischer Darmstörungen mit
832 Colibakterien. Med Klin 29–30.
- 833 19. Mathieu A, Dion M, Deng L, Tremblay D, Moncaut E, Shah SA, Stokholm J, Krogfelt
834 KA, Schjørring S, Bisgaard H, Nielsen DS, Moineau S, Petit MA. 2020. Virulent
835 coliphages in 1-year-old children fecal samples are fewer, but more infectious than
836 temperate coliphages. Nat Commun 11:1–12.
- 837 20. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment
838 search tool. J Mol Biol 215:403–410.
- 839 21. Adriaenssens EM, Rodney Brister J. 2017. How to name and classify your phage: An
840 informal guide. Viruses 9:1–9.
- 841 22. International Committee on Taxonomy of Viruses. 2020. Master Species Lists. Retrieved
842 from <https://talk.ictvonline.org/files/master-species-lists/>.
- 843 23. Bin Jang H, Bolduc B, Zablocki O, Kuhn JH, Roux S, Adriaenssens EM, Brister JR,
844 Kropinski AM, Krupovic M, Lavigne R, Turner D, Sullivan MB. 2019. Taxonomic
845 assignment of uncultivated prokaryotic virus genomes is enabled by gene-sharing
846 networks. Nat Biotechnol 37:632–639.

24. Neal BL, Brown PK, Reeves PR. 1993. Use of Salmonella phage P22 for transduction in *Escherichia coli*. *J Bacteriol* 175:7115–7118.
25. Jakhetia R, Marri A, Ståhle J, Widmalm G, Verma NK. 2014. Serotype-conversion in *Shigella flexneri*: Identification of a novel bacteriophage, Sf101, from a serotype 7a strain. *BMC Genomics* 15.
26. Clark AJ, Inwood W, Cloutier T, Dhillon TS. 2001. Nucleotide sequence of coliphage HK620 and the evolution of lambdoid phages. *J Mol Biol* 311:657–679.
27. Nakata K, Koh MM, Tsuchido T, Matsumura Y. 2010. All genomic mutations in the antimicrobial surfactant-resistant mutant, *Escherichia coli* OW66, are involved in cell resistance to surfactant. *Appl Microbiol Biotechnol* 87:1895–1905.
28. McFall E, Heincz M. 1983. Identification and control of synthesis of the *dsdC* activator protein. *J Bacteriol* 153:872–877.
29. Moritz RL, Welch RA. 2006. The *Escherichia coli* *argW-dsdCXA* genetic island is highly variable, and *E. coli* K1 strains commonly possess two copies of *dsdCXA*. *J Clin Microbiol* 44:4038–4048.
30. Reiter W-D, Palm P, Yeats S. 1989. Transfer RNA genes frequently serve as integration sites for prokaryotic genetic elements. *Nucleic Acids Res* 17.
31. Casjens S, Winn-Stapley DA, Gilcrease EB, Morona R, Kühlewein C, Chua JEH, Manning PA, Inwood W, Clark AJ. 2004. The chromosome of *Shigella flexneri* bacteriophage Sf6: Complete nucleotide sequence, genetic mosaicism, and DNA packaging. *J Mol Biol* 339:379–394.

- 868 32. Panis G, Méjean V, Ansaldi M. 2007. Control and regulation of KplE1 prophage site-
869 specific recombination: A new recombination module analyzed. *J Biol Chem* 282:21798–
870 21809.
- 871 33. Abedon ST. 2018. Detection of Bacteriophages: Phage Plaques, p. 1–32. *In*
872 *Bacteriophages*.
- 873 34. Silva JB, Storms Z, Sauvageau D. 2016. Host receptors for bacteriophage adsorption.
874 *FEMS Microbiol Lett* 363:1–11.
- 875 35. Dhillon TS, Poon APW, Chan D, Clark AJ. 1998. General transducing phages like
876 *Salmonella* phage P22 isolated using a smooth strain of *Escherichia coli* as host. *FEMS*
877 *Microbiol Lett* 161:129–133.
- 878 36. Samuel G, Reeves P. 2003. Biosynthesis of O-antigens: Genes and pathways involved in
879 nucleotide sugar precursor synthesis and O-antigen assembly. *Carbohydr Res* 338:2503–
880 2519.
- 881 37. Heinrichs DE, Monteiro MA, Perry MB, Whitfield C. 1998. The assembly system for the
882 lipopolysaccharide R2 core-type of *Escherichia coli* is a hybrid of those found in
883 *Escherichia coli* K-12 and *Salmonella enterica*. Structure and function of the R2 WaaK
884 and WaaL homologs. *J Biol Chem* 273:8849–8859.
- 885 38. Gronow S, Brabetz W, Brade H. 2000. Comparative functional characterization in vitro of
886 heptosyltransferase I (WaaC) and II (WaaF) from *Escherichia coli*. *Eur J Biochem*
887 267:6602–6611.
- 888 39. Wang Z, Wang J, Ren G, Li Y, Wang X. 2016. Deletion of the genes *waaC*, *waaF*, or

889 waaG in *Escherichia coli* W3110 disables the flagella biosynthesis. *J Basic Microbiol*
890 56:1021–1035.

891 40. Verma NK, Brandt JM, Verma DJ, Lindberg AA. 1991. Molecular characterization of the
892 O-acetyl transferase gene of converting bacteriophage SF6 that adds group antigen 6 to
893 *Shigella flexneri*. *Mol Microbiol* 5:71–75.

894 41. Vander Byl C, Kropinski AM. 2000. Sequence of the genome of *Salmonella*
895 bacteriophage P22. *J Bacteriol* 182:6472–6481.

896 42. Amann E, Ochs B, Abel KJ. 1988. Tightly regulated tac promoter vectors useful for the
897 expression of unfused and fused proteins in *Escherichia coli*. *Gene* 69:301–315.

898 43. Newton GJ, Daniels C, Burrows LL, Kropinski AM, Clarke AJ, Lam JS. 2001. Three-
899 component-mediated serotype conversion in *Pseudomonas aeruginosa* by bacteriophage
900 D3. *Mol Microbiol* 39:1237–1247.

901 44. Norrander J, Kempe T, Messing J. 1983. Construction of improved M13 vectors using
902 oligodeoxynucleotide-directed mutagenesis. *Gene* 26:101–106.

903 45. Ferrières L, Hémerly G, Nham T, Guérout AM, Mazel D, Beloin C, Ghigo JM. 2010.
904 Silent mischief: Bacteriophage Mu insertions contaminate products of *Escherichia coli*
905 random mutagenesis performed using suicidal transposon delivery plasmids mobilized by
906 broad-host-range RP4 conjugative machinery. *J Bacteriol* 192:6418–6427.

907 46. Edwards RA, Keller LH, Schifferli DM. 1998. Improved allelic exchange vectors and
908 their use to analyze 987P fimbria gene expression. *Gene* 207:149–157.

909 47. Broecker NK, Barbirz S. 2017. Not a barrier but a key: How bacteriophages exploit host's

910 O-antigen as an essential receptor to initiate infection. *Mol Microbiol* 105:353–357.

911 48. Millette M, Nguyen A, Amine KM, Lacroix M. 2013. Gastrointestinal survival of bacteria
912 in commercial probiotic products. *Int J Probiotics Prebiotics* 8:149–156.

913 49. Dodd IB, Shearwin KE, Egan JB. 2005. Revisited gene regulation in bacteriophage λ .
914 *Curr Opin Genet Dev* 15:145–152.

915 50. Price NL, Raivio TL. 2009. Characterization of the Cpx regulon in *Escherichia coli* strain
916 MC4100. *J Bacteriol* 191:1798–1815.

917 51. Wang X, Kim Y, Ma Q, Hong SH, Pokusaeva K, Sturino JM, Wood TK. 2010. Cryptic
918 prophages help bacteria cope with adverse environments. *Nat Commun* 1:147.

919 52. De Paepe M, Tournier L, Moncaut E, Son O, Langella P, Petit MA. 2016. Carriage of λ
920 Latent Virus Is Costly for Its Bacterial Host due to Frequent Reactivation in Monoxenic
921 Mouse Intestine. *PLoS Genet* 12:1–20.

922 53. Oh JH, Lin XB, Zhang S, Tollenaar SL, Özcam M, Dunphy C, Walter J, van Pijkeren JP.
923 2020. Prophages in *Lactobacillus reuteri* are associated with fitness trade-offs but can
924 increase competitiveness in the gut ecosystem. *Appl Environ Microbiol* 86:1–22.

925 54. GraphPad Software. GraphPad Prism. Retrieved from www.graphpad.com.

926 55. Peters DL, Mccutcheon JG, Stothard P, Dennis JJ. 2019. Novel *Stenotrophomonas*
927 maltophilia temperate phage DLP4 is capable of lysogenic conversion. *BMC Genomics*
928 20.

929 56. Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S,
930 Rueden C, Saalfeld S, Schmid B, Tinevez JY, White DJ, Hartenstein V, Eliceiri K,

- Tomancak P, Cardona A. 2012. Fiji: An open-source platform for biological-image analysis. *Nat Methods* 9:676–682.
57. Center for Phage Technology. 2018. Protocol for Phage DNA Extraction with Phenol:Chloroform. Retrieved from <https://cpt.tamu.edu/phage-links/phage-protocols>.
58. Afgan E, Baker D, Batut B, Van Den Beek M, Bouvier D, Ech M, Chilton J, Clements D, Coraor N, Grüning BA, Guerler A, Hillman-Jackson J, Hiltmann S, Jalili V, Rasche H, Soranzo N, Goecks J, Taylor J, Nekrutenko A, Blankenberg D. 2018. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Res* 46:W537–W544.
59. Babraham Bioinformatics. 2012. Trim Galore! A wrapper tool around Cutadapt and FastQC to consistently apply quality and adapter trimming to FastQ files. Retrieved from https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/.
60. Babraham Bioinformatics. 2010. FastQC, A quality control tool for high throughput sequence data. Retrieved from <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.
61. Ewels P, Magnusson M, Lundin S, Käller M. 2016. MultiQC: Summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* 32:3047–3048.
62. Wick RR, Judd LM, Gorrie CL, Holt KE. 2017. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput Biol* 13:1–16.
63. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin A V., Sirotkin A V., Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: A new genome assembly algorithm and its

applications to single-cell sequencing. *J Comput Biol* 19:455–477.

64. Mikheenko A, Prjibelski A, Saveliev V, Antipov D, Gurevich A. 2018. Versatile genome assembly evaluation with QUAST-LG. *Bioinformatics* 34:i142–i150.

65. Seemann T. 2014. Prokka: Rapid prokaryotic genome annotation. *Bioinformatics* 30:2068–2069.

66. Cuccuru G, Orsini M, Pinna A, Sbardellati A, Soranzo N, Travaglione A, Uva P, Zanetti G, Fotia G. 2014. Orione, a web-based framework for NGS analysis in microbiology. *Bioinformatics* 30:1928–1929.

67. Salisbury A, Tsourkas PK. 2019. A method for improving the accuracy and efficiency of bacteriophage genome annotation. *Int J Mol Sci* 20.

68. Besemer J. 2001. GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. *Nucleic Acids Res* 29:2607–2618.

69. Delcher A, Bratke K, Powers E, Salzberg S. 2007. Identifying bacterial genes and endosymbiont DNA with Glimmer. *Bioinformatics* 23:673–679.

70. Marchler-Bauer A, Lu S, Anderson JB, Chitsaz F, Derbyshire MK, DeWeese-Scott C, Fong JH, Geer LY, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Jackson JD, Ke Z, Lanczycki CJ, Lu F, Marchler GH, Mullokandov M, Omelchenko M V., Robertson CL, Song JS, Thanki N, Yamashita RA, Zhang D, Zhang N, Zheng C, Bryant SH. 2011. CDD: A Conserved Domain Database for the functional annotation of proteins. *Nucleic Acids Res* 39:225–229.

71. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. 2003. Cytoscape: A Software Environment for Integrated Models. *Genome Res* 13:2498–2504.
72. Hitchcock PJ. 1984. Analyses of gonococcal lipopolysaccharide in whole-cell lysates by sodium dodecyl sulfate-polyacrylamide gel electrophoresis: Stable association of lipopolysaccharide with the major outer membrane protein (protein I) of *Neisseria gonorrhoeae*. *Infect Immun* 46:202–212.
73. Kulikov EE, Golomidova AK, Prokhorov NS, Ivanov PA, Letarov A V. 2019. High-throughput LPS profiling as a tool for revealing of bacteriophage infection strategies. *Sci Rep* 9.
74. Tsai C, Frasch C. 1982. Silver Stain for Detecting Lipopolysaccharides Polyacrylamide Gels. *Anal Biochem* 19:115–119.
75. Davis MR, Goldberg JB. 2012. Purification and visualization of lipopolysaccharide from gram-negative bacteria by hot aqueous-phenol extraction. *J Vis Exp* 1.
76. Deatherage DE, Barrick JE. 2014. Identification of mutations in laboratory evolved microbes from next-generation sequencing data using breseq. *Methods Mol Biol* https://doi.org/10.1007/978-1-4939-0554-6_12.
77. Seemann T. 2015. snippy: fast bacterial variant calling from NGS reads. Retrieved from <https://github.com/tseemann/snippy>.
78. Sambrook J, Russell DW. 2001. *Molecular cloning: a laboratory manual* 3rd edition. Coldspring Harbour Laboratory Press, UK.

994 79. Silhavy TJ, Berman ML, Enquist LW. 1984. Experiments with gene fusions. Cold Spring
995 Harb Lab.

996 80. Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, Datsenko KA, Tomita M,
997 Wanner BL, Mori H. 2006. Construction of Escherichia coli K-12 in-frame, single-gene
998 knockout mutants: The Keio collection. Mol Syst Biol
999 <https://doi.org/10.1038/msb4100050>.

1000 81. Datsenko KA, Wanner BL. 2000. One-step inactivation of chromosomal genes in
1001 Escherichia coli K-12 using PCR products. Proc Natl Acad Sci U S A 97:6640–6645.

1002 82. Vogt SL, Scholz R, Peng Y, Guest RL, Scott NE, Woodward SE, Foster LJ, Raivio TL,
1003 Finlay BB. 2019. Characterization of the Citrobacter rodentium Cpx regulon and its role
1004 in host infection. Mol Microbiol 111:700–716.

1005 83. Lederberg EM, Lederberg J. 1953. Genetic Studies of Lysogenicity in Escherichia coli.
1006 Genetics 38:51–64.

1007 84. Wintersinger JA, Wasmuth JD. 2015. Kablammo: An interactive, web-based BLAST
1008 results visualizer. Bioinformatics 31:1305–1306.

1009 85. Kropinski AM. 2009. Measurement of the rate of attachment of bacteriophage to cells., p.
1010 151–155. In Bacteriophages, Methods and Protocols, Volume 1: Isolation,
1011 Characterization, and Interactions.

1012 86. Guyer MS, Reed RR, Steitz JA, Low KB. 1981. Identification of a sex-factor-affinity site
1013 in E. coli as gamma delta. Cold Spring Harb Symp Quant Biol 45:135–140.

1014 87. Casadaban MJ. 1976. Transposition and fusion of the lac genes to selected promoters in

Escherichia coli using bacteriophage lambda and Mu. J Mol Biol 104:541–555.

88. Bachmann BJ. 1972. Pedigrees of some mutant strains of Escherichia coli K-12. Bacteriol Rev 36:525–557.

89. Schauer DB, Falkow S. 1993. The eae gene of Citrobacter freundii biotype 4280 is necessary for colonization in transmissible murine colonic hyperplasia. Infect Immun 61:4654–4661.

Figure Legends

Figure 1. PHASTER analysis of MP1 and transmission electron microscopy image of Kapi1. (A) The genome of MP1 (accession JEMI01000000) (11) was analyzed for putative prophage regions using PHASTER (12, 13). Prophage regions identified by PHASTER are shown, including the most common phage from the NCBI viral database, and their corresponding accession numbers. Intact, questionable, and incomplete scores were assigned by PHASTER. (B) Kapi1 phage lysate was stained with 4% uranyl acetate on a copper grid, and viewed by transmission electron microscopy at 140,000x magnification.

Figure 2. Genome of Kapi1, and alignment with the most closely related phage, Sf101.

Above, layout of the Kapi1 genome, with ORFs color coded as follows: lysis module in blue, structural/morphogenesis in pink, DNA replication/repair/regulation in green, tRNAs in yellow, and hypothetical proteins and proteins of unknown function in grey. Below, the alignment of the

Kapi1 genome against the genome of phage Sf101 is visualized with Kablammo (84); darker lines represent higher sequence homology between the two phage. Sf101 seroconversion protein (*gp16* O-acyltransferase B) is indicated in purple.

Figure 3. The integration site of Kapi1. (A) The structure of Kapi1 integrated into the host chromosome; host ORFs are indicated in orange, and phage ORFs are indicated in green and grey, in between the *attL* and *attR* sites. Nucleotide sequences are provided for both the left and right host-prophage chromosome junctions, including 20 bp upstream and downstream of the bolded and underlined *att* sites. Grey shading on the junction_right sequence indicates the location of the Kapi1 *hyp25* gene, and orange shading indicates the host *argW* gene. (B) Kapi1 phage lysate was serially diluted and spotted onto soft-agar overlays of either WT KP7 or KP7 $\Delta attB::kan$.

Figure 4. Kapi1 uses the O-antigen as a receptor and modifies its structure. (A) LPS extracted by proteinase K digestion was run on SDS-PAGE and silver stained; spontaneous Kapi1-resistnat mutants in lanes 3 and 4 correspond to KP61 and KP62, respectively. (B) Schematic representing the truncated *wzy* polymerase from KP61 and KP62 compared with the full-length WT protein. (C) Kapi1 phage lysate was serially diluted and spotted onto soft-agar overlays with a variety of mutants in the KP7 background. (D) Phenol-extracted LPS was run on SDS-PAGE and silver stained; in lane 2, KP7 LPS was treated with Kapi1 phage prior to running on SDS-PAGE.

Figure 5. Stability of Kapi1 lysogens grown in different media. (A) Cultures of MP13 (Kapi1 lysogen) were grown in LB, 50% LB 50% simulated gastric fluid (SGF) (48), 50% LB 50% simulated intestinal fluid (SIF) (48), or LB supplemented with 0.5 ng/uL mitomycin C for 24 hours. After 24 hours the number of cells were enumerated by spotting on LB plates, and the number of phages were enumerated by spotting on soft-agar overlays prepared with KP121. PFU/mL was divided by CFU/mL to obtain PFU/CFU; the values represent the average of three biological replicates, and error bars represent the standard deviation. One outlier value was excluded for the mitomycin C treatment. A one-way ANOVA was performed with a Tukey's post-hoc test on log-transformed data. (B) MP13 was grown in the same media as above for 24 hours, then sub-cultured into fresh media and grown for another 24 hours. After 48 hours total incubation, cells were plated onto LB agar, then replica plated onto a second LB agar plate spread with KP7 to screen for Kapi1 lysogens (Figure S5). The number of non-lysogenic colonies was divided by the total number of colonies to obtain % loss of Kapi1; values represent the average of three biological replicates, and error bars represent the standard deviation. A one-way ANOVA was performed with a Tukey post-hoc test on arcsin-transformed data. (C) Luminescent reporter assay; the activity of Kapi1 *CI-lux* was monitored in Kapi1 lysogens grown in the same media as above. The average luminescence (counts per second, CPS) normalized to bacterial optical density (OD^{600}) is reported for three biological replicates, and error bars represent the standard deviation. (D) *CI-lux* activity after 24 hours of growth. The average normalized luminescence is reported for three biological replicates, and error bars represent the standard deviation. A one-way ANOVA was performed with a Tukey post-hoc test.

Figure 6. Kapi1 lysogens have a competitive advantage over their non-lysogenic counterparts. (A) *In vitro* competitive assay; KP7 and MP13 were co-cultured in either LB or SIF. Each strain was enumerated every 24 hours, after which co-cultures were sub-cultured into fresh media and grown for an additional 24 hours. The relative proportion of each strain was calculated by dividing KP7 or MP13 CFU/mL by the total CFU/mL. The average of 3 biological replicates is reported, with the standard deviation as error bars. A t-test assuming unequal variance was performed on arcsin-transformed data at t=48 hr and found no statistical significance between the two treatments ($p > 0.05$). (B) *In vivo* colonization; coliform-free mice were gavaged with either KP7 or MP13, and log (CFU/g feces) was enumerated by plating on MacConkey media after 14 days. The average of 4 mice is reported, with standard deviation as error bars. A t-test assuming unequal variance was performed and found no statistical significance between the two treatments ($p > 0.05$). (C) *In vivo* competitive colonization; coliform-free mice were gavaged with a mixture of KP7 and MP13, and the relative proportion of each strain was enumerated after 14 days. The average of 10 mice is reported, with standard deviation as error bars.

Figure S1. PCR of DNase-treated lysates for putative prophages identified by PHASTER. Primers were designed to amplify each of the putative prophage regions on the MP13 genome identified by PHASTER (12, 13) (I = Intact_1, Q = Questionable, IC = Incomplete). Cultures of KP7 and MP13 were grown for 24 hours with 0.5 ng/ μ L mitomycin C, then culture supernatants were collected, filter sterilized (0.45 μ m), and treated twice with 0.01 mg/mL DNase A to remove genomic DNA. DNase A was inactivated with 0.5 M EDTA at room temperature for 3 minutes, then 5 μ L of the treated lysates were used for PCR. As a positive control, 5 μ L of MP13

culture was also used for PCR to ensure that the prophage primers worked correctly. To control for gDNA contamination, a set of primers amplifying chromosomal *nuoA* (N) was included.

Figure S2. vContact2 (23) clustering of Kapi1-related viruses, visualized with CytoScape (71). The entire first-neighbors network of Kapi1 is shown. An edge-weighted, spring-embedded model was used so that the distance between viral nodes, and the darkness and thickness of edges connecting those nodes corresponds to their sequence similarity and relatedness. Viral clusters are color coded.

Figure S3. Amino acid alignment of WT and truncated *wzy* polymerases. The amino acids of truncated *wzy* polymerase from Kapi1-resistant LPS mutants KP61 and KP62 were aligned with the WT protein from KP7 using CLC Genomics Workbench 21 (<https://digitalinsights.qiagen.com/>).

Figure S4. Cell growth and phage adsorption rates, additional controls for PFU/CFU experiments. (A) Growth rates of KP7 and MP13 in LB, calculated as the slope of exponential phase. Overnight cultures were adjusted to OD⁶⁰⁰ 1.0 in PBS, then sub-cultured 1:100 into a sterile 96-well plate and incubated at 37 C with agitation in the Epoch2 microplate reader (Biotek, USA); OD⁶⁰⁰ was measured every 30 minutes for 16 hours. the average of 3 biological replicates and 2 technical replicates (n=6) is reported, with the standard deviation as error bars. A t-test assuming unequal variance was performed (p < 0.001). (B) The adsorption rate of Kapi1 to KP7 in different media. Adsorption rate was determined experimentally as previously described

(85), the average of 3 experiments is reported, with the standard deviation as error bars. A one-way ANOVA was performed ($p > 0.05$). (C) Growth rates of MP13 in LB, SGF, and SIF, as above. A one-way ANOVA was performed with Tukey's post-hoc test ($p < 0.001$). (D) The data shown is identical to the data reported in Figure 5A, with the inclusion of 50% LB / 50% PBS and 50% LB / 50% distilled water controls. A one-way ANOVA was performed with Tukey's post-hoc test ($p < 0.001$). One outlier in the 50% LB / 50% distilled water treatment was excluded.

Figure S5. Experimental verification of replica plating lysogeny screening method. (A) 30 colonies of each phenotype (L – lysogens, NL – non-lysogens) were used for PCR to check for both free phage (FP, using primers for Kapi1 coat protein) and prophage (PP, using primers spanning the genome-prophage junction). All colonies were grown up overnight, then checked for superinfection immunity to Kapi1 by spotting 5 μ L high-titre Kapi1 lysate onto soft-agar overlays of each strain. Colonies were also checked for production of Kapi1 by filter-sterilizing culture supernatants and spotting 5 μ L onto KP7 soft-agar overlays. 30/30 colonies scored as lysogens were positive for free phage and prophage by PCR, and positive for superinfection immunity and Kapi1 production by spotting. 19/30 colonies scored as non-lysogens were negative for free phage and prophage by PCR, and negative for superinfection immunity and Kapi1 production by spotting. 8/30 colonies scored as non-lysogens were positive for free phage, but negative for prophage by PCR. As expected, these colonies became lysogenic after growing them up overnight, as indicated by positive superinfection immunity and Kapi1 production by spotting. This result indicates that this method is extremely sensitive for capturing lysogens; these nine colonies were most likely recently infected, or may have Kapi1 virions bound to the

cell surface, but not yet integrated into the genome, and therefore no productive phage particles are released onto the lawn of KP7 producing the halo phenotype. 3/30 colonies scored as non-lysogens were negative for free phage and prophage by PCR, but became lysogenic after overnight incubation, as indicated by positive superinfection and Kapi1 production by spotting. This result was unexpected, but we believe that these three colonies likely already had phage bound (like the previous group of 9), but the number of phage may have been too low to be detected by PCR, since a genome prep wasn't performed, colony suspensions were used directly for PCR. (B) Representative phenotypes: black arrows indicate non-lysogenic colonies without a zone of clearing, white arrows indicate Kapi1 lysogens producing a zone of clearing on an KP7 lawn.

Table S1. Bacterial strains used in this study.

Table S2. Nucleotide primers used in this study^a

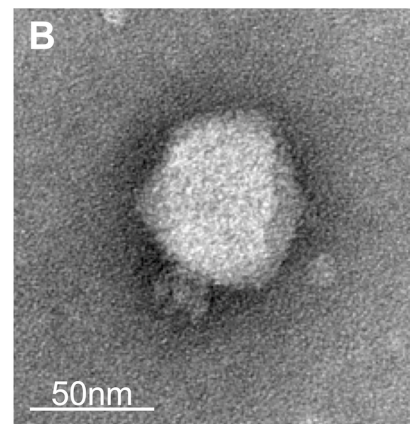
^a All primers were designed in this study

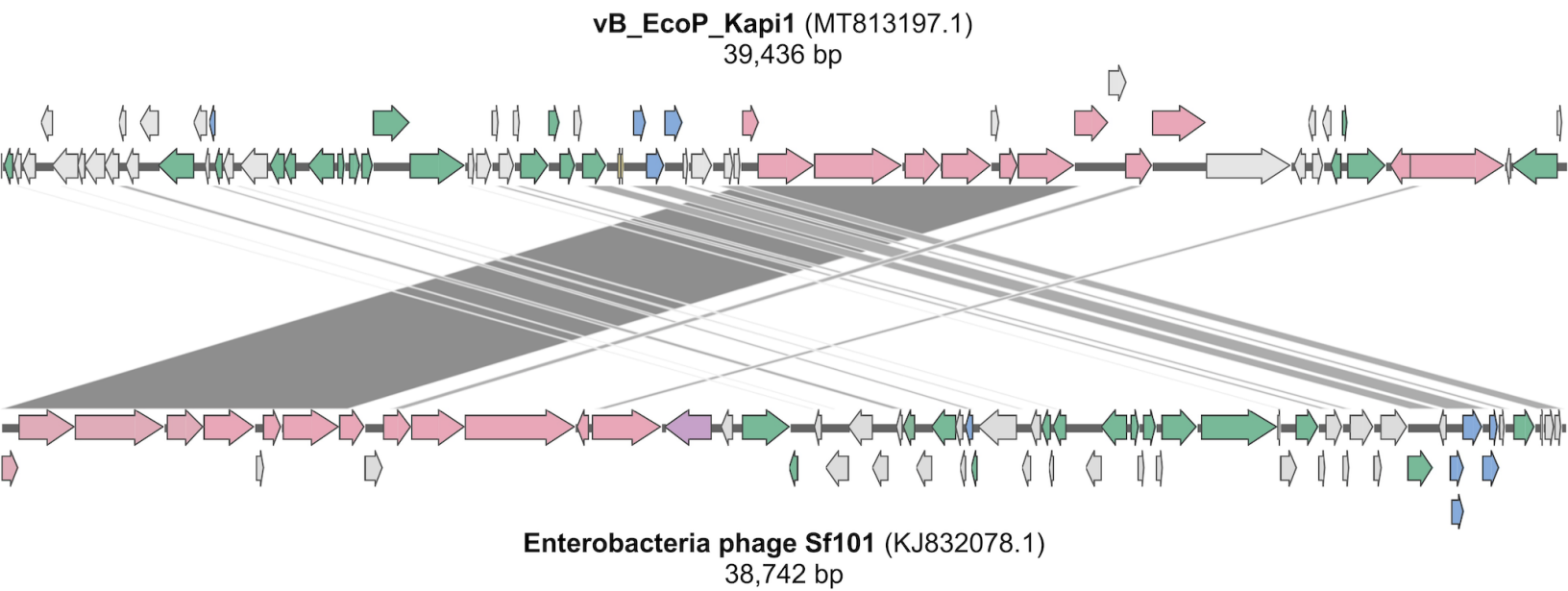
Table S3. Genome annotation of bacteriophage Kapi1.

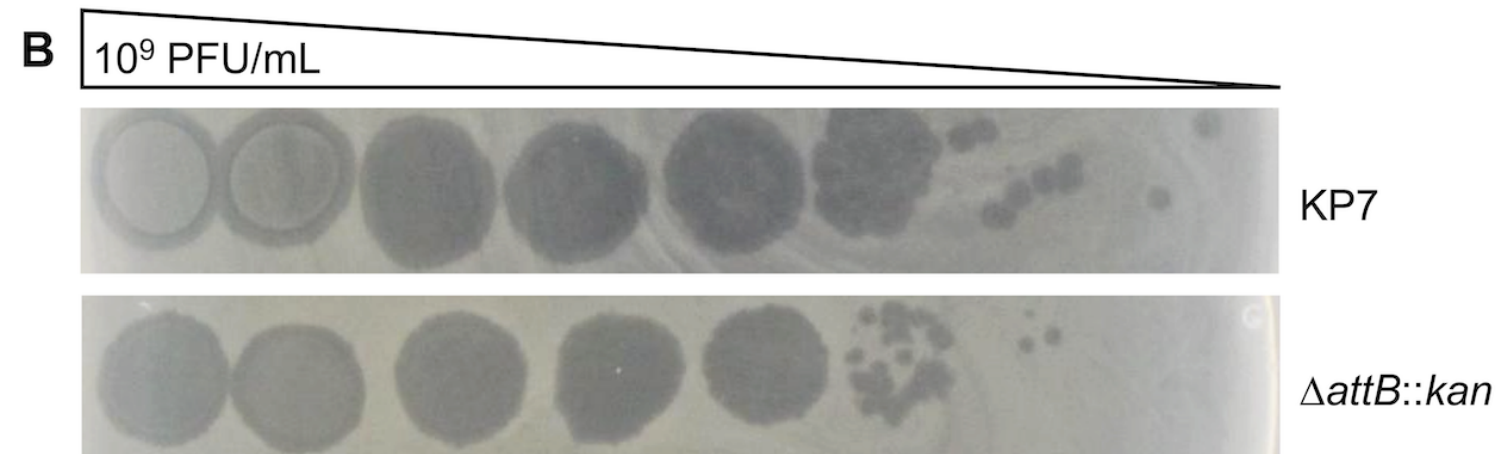
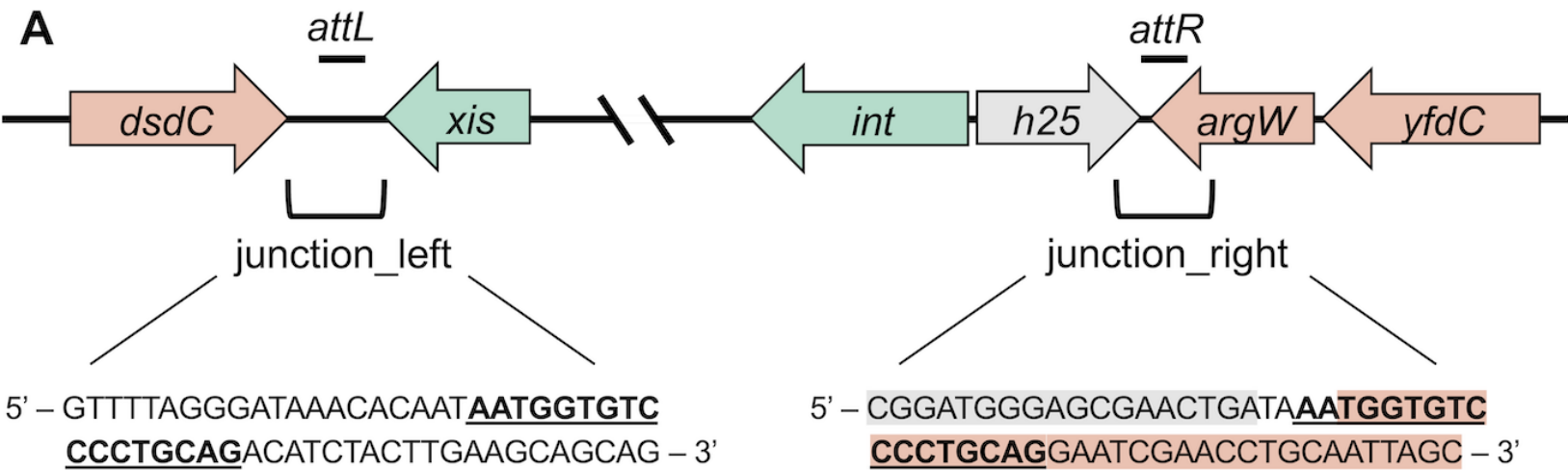
* putative O-antigen modification genes

A

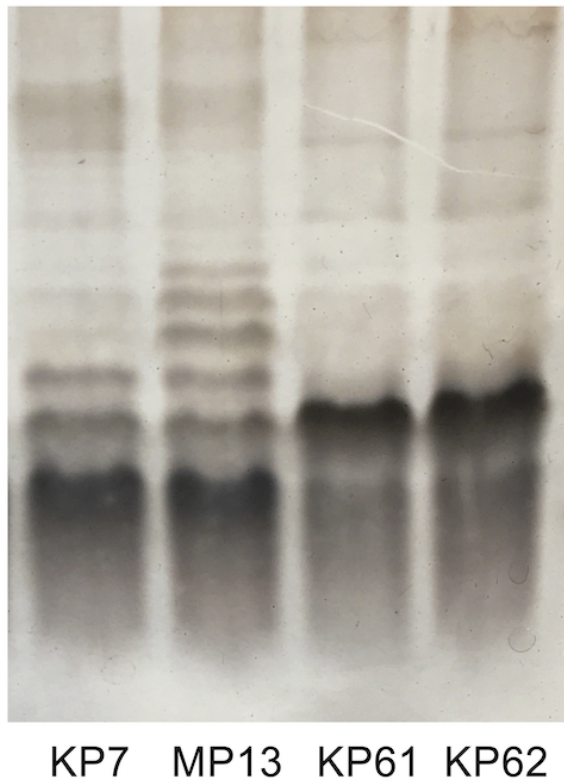
Prophage Region	Most Common Phage	Accession
Intact_1	Enterobacteria phage Sf101	NC_027398
Questionable_1	Enterobacteria phage DE3	NC_042057
Questionable_2	Stx2-converting phage Stx2a-F451	NC_049924
Incomplete_1	Bacteriophage APSE-2	NC_011551
Incomplete_2	Escherichia phage 500456-1	NC_049342
Incomplete_3	Bacillus phage G	NC_023719



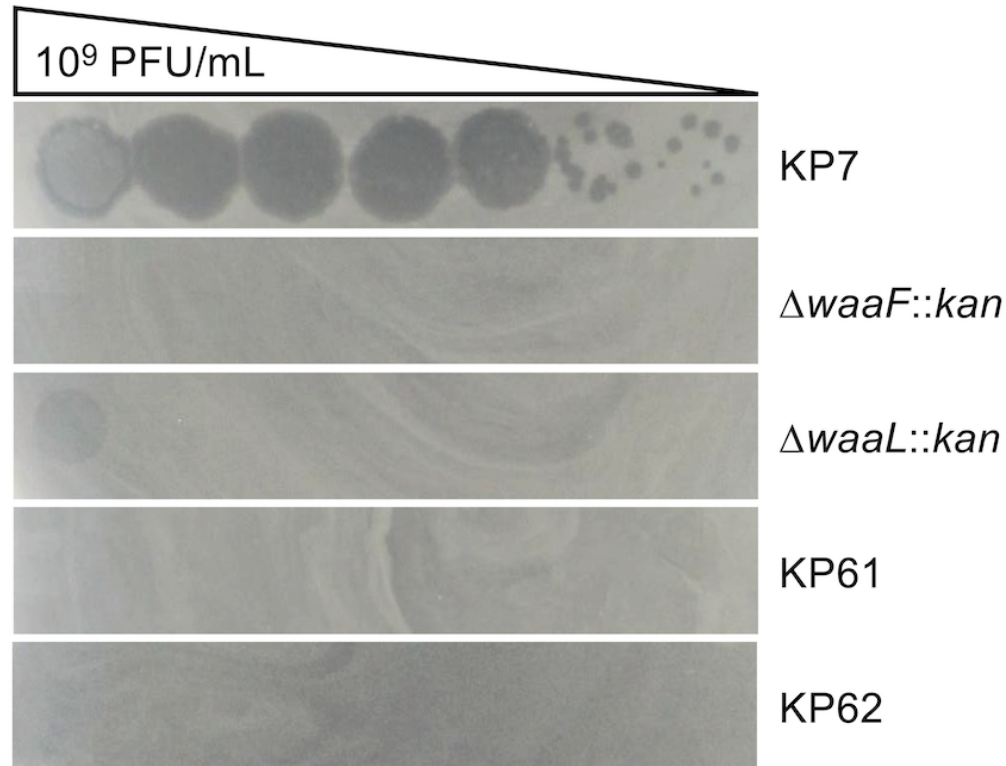




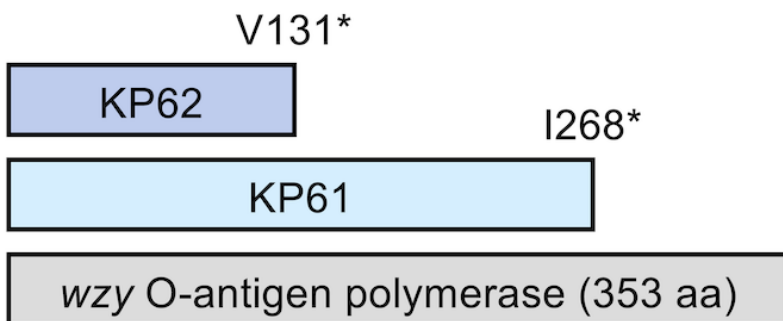
A



C



B



D

