

1 Isolation and characterization of a novel bacteriophage, Kapi1, capable of O-antigen modification
2 in commensal *Escherichia coli*.

3

4 Kat Pick^a and Tracy Raivio^a

5

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

7

8 Running Title: Characterization of novel bacteriophage Kapi1

9

10 # Address correspondence to Tracy Raivio, traivio@ualberta.ca

11

12 Abstract word count: 152 (importance: 125)

13

14 Text word count: 4958

15

16

17

18

19

20 **Abstract**

21 In this study, we describe the isolation and characterization of novel bacteriophage Kapi1
22 (vB_EcoP_Kapi1) isolated from a strain of commensal *Escherichia coli* inhabiting the
23 gastrointestinal tract of healthy mice. We show that Kapi1 is a temperate phage integrated into
24 tRNA *argW* of strain MP1 and describe its genome annotation and structure. Kapi1 shows limited
25 homology to other characterized prophages but is most similar to the phages of *Shigella flexneri*,
26 and clusters taxonomically with P22-like phages. Investigation of the lifestyle of Kapi1 shows that
27 this phage displays unstable lysogeny and influences the growth of its host. The receptor for Kapi1
28 is the lipopolysaccharide O-antigen, and we further show that Kapi1 alters the structure of its hosts
29 O-antigen in multiple ways. We hope to use MP1 and Kapi1 as a model system to explore
30 molecular mechanisms of mammalian colonization by *E. coli* and ask what the role(s) of prophages
31 in this context might be.

32

33 **Importance**

34 Although research exploring the microbiome has exploded in recent years, our
35 understanding of the viral component of the microbiome is lagging far behind our understanding
36 of the bacterial component. The vast majority of intestinal bacteria carry prophages integrated into
37 their chromosomes, but most of these bacteriophages remain uncharacterized and unexplored.
38 Here, we isolate and characterize a novel temperate bacteriophage infecting a commensal strain of
39 *Escherichia coli*. We aim to explore the interactions between bacteriophages and their hosts in the
40 context of the gastrointestinal tract, asking what role(s) temperate bacteriophage may play in
41 growth and survival of bacteria in the gastrointestinal tract. Understanding the fundamental

42 biology of commensal bacteria in the gastrointestinal tract can inform the development of novel
43 antimicrobial or probiotic strategies for intestinal infections.

44

45 **Introduction**

46 *Escherichia coli* is a Gram-negative bacterium normally inhabiting the lower
47 gastrointestinal tract of humans and other warm-blooded animals (1). Despite being one of the
48 most widely studied prokaryotic model organisms, there remains an immense complexity to the
49 lifestyle of *E. coli* that we are only beginning to appreciate; one of these layers of complexity is
50 the interactions between *E. coli* and the bacteriophages that infect it. Bacteriophage (or simply
51 phage) exhibit two main lifestyles; lytic phage infect and immediately begin replicating within
52 their host, eventually causing cell lysis and releasing progeny phages. Temperate phage integrate
53 into the genome of their host where they reside as mostly dormant prophages, replicating with the
54 host chromosome and being disseminated into daughter cells. Once the host cell experiences stress
55 such as DNA damage or starvation, the prophage excises from the chromosome and enters the
56 lytic cycle to ensure its' own survival (2). Temperate phage have been gaining attention as we
57 begin to appreciate their abundance; it has been estimated that approximately half of all sequenced
58 bacterial genomes contain intact prophage, and even more contain prophage elements (3).
59 Interestingly, the abundance of temperate phage residing in the commensal gut microbiome of
60 mice appears to be even higher (4), indicating that temperate phage may play a role in bacterial
61 community dynamics during colonization. Indeed, many recent studies and reviews have
62 highlighted the importance of phages in the microbiome community (5–8).

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

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

82

83 **Results and Discussion**

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

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

107 Because our cultures of MP1 and MP13 containing phage appeared healthy, we
108 hypothesized that the phage in these cultures may be lysogenic, as a lytic phage would be more
109 likely to lyse the cultures resulting in a visible reduction in cell density and poor growth. Analysis
110 of the previously published genome sequence for MP1 (11) using the prophage identification tool
111 PHASTER (12, 13) revealed six putative prophages integrated into the chromosome of MP1
112 (Figure 1A). Of these, only one prophage was scored as intact by PHASTER (12, 13) (Figure 1A);
113 we hypothesized that this prophage was the most likely candidate for the phage plaques we had
114 observed because of the completeness of the prophage sequence. To confirm this, we performed
115 PCR on colonies of MP1, MP7 and MP13, as well as phage lysates prepared from MP1 and MP13
116 with three primer pairs targeting the coat, portal, and tail proteins within the Intact_1 prophage
117 region identified by PHASTER. Bands were consistently observed for MP1 and MP13 colonies
118 and phage lysates and were consistently absent for MP7 colonies for all primer pairs (data not
119 shown). This indicated that the phage present in MP1 and MP13 cultures likely corresponds to the
120 Sf101-like Intact_1 prophage region identified by PHASTER in the MP1 genome. Despite several
121 attempts to induce phage from MP7 using either DNA damaging agents or cellular starvation,
122 phage could never be isolated from MP7. Since Lasaro et al. (2014) previously showed that MP7
123 and MP13 had equal competitive indices *in vivo*, we decided to investigate this further. Upon
124 testing the original stock of MP7 isolated by Lasaro et al. (2014), we found that this strain does in
125 fact contain the Intact_1 prophage region (data not shown), and that only our stock of MP7 lacks
126 the Intact_1 prophage region. These findings are important as they demonstrate that any
127 competitive advantage that would have been provided by carrying the Intact_1 prophage did not
128 play a role in the findings of Lasaro et al. (2014), since both MP7 and MP13 contain the prophage.
129 We propose that our stock of MP7 which lacks the Intact_1 prophage region should be renamed

130 to KP7, in order to avoid confusion with the original MP7 stock which does contain the Intact_1
131 prophage region.

132 Transmission electron microscopy (Figure 1B) of phage lysates collected from MP1 and
133 MP13 revealed phage particles with a mean capsid diameter of 70.10 ± 2.92 nm and tail length
134 15.37 ± 1.45 nm placing this phage in the family *Podoviridae* and order *Caudovirale*. We screened
135 11 strains of *E. coli* (Top10, MG1655, TJ-LM, TJ-WM, TJ-LR, MC4100, W3110, BW25113, J96,
136 E2348/69, Nissile 1917) and *Citrobacter rodentium* DBS100 for susceptibility to the phage (see
137 Table S1). All strains were completely resistant to infection; from our strain collection, KP7 is the
138 only strain that this phage can infect. This preliminary analysis suggests that the newly isolated
139 phage has a relatively narrow host range. This phage forms diffuse plaques on KP7 with an average
140 plaque diameter of 2.0 ± 0.22 mm after overnight incubation at 37 °C. Although the morphology
141 of phage particles on TEM and plaques on soft agar overlays was consistent, to confirm that the
142 Intact_1 prophage region is the only prophage in MP1 capable of active excision and lytic
143 replication, we performed PCR on DNase-treated phage lysates using primers corresponding to
144 each putative prophage region identified by PHASTER, and a *nuoA* primer pair to control for
145 genomic DNA contamination. No bands were observed in KP7 lysates, and only the band
146 corresponding to Intact_1 prophage was observed in the MP13 lysate (Figure S1).

147

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

149 Although the whole genome sequence for MP1 has already been published (11), to be
150 thorough and ensure that the isolated phage was truly a prophage and not introduced by
151 contamination from our laboratory, whole genome sequencing was performed on our stocks of
152 MP1, KP7, and MP13. We also aimed to find the integration site and characterize the genome of

153 the isolated phage. As anticipated, the 39 kb Sf101-like region was present in the genomes of
154 MP1 and MP13, and absent from the genome of KP7. Unfortunately, the phage genome was
155 assembled into its own linear contig, not showing where it may be integrated into the host
156 chromosome. This was observed in both the original MP1 sequence (11) and in our resequencing
157 attempt; it is likely that the phage genome was in its circular form (ie. excised from the host
158 chromosome) since we extracted DNA from late stationary phase cultures. Upon closer analysis
159 of the original MP1 sequence (11), it appears that the ends of contig NZ_JEMI01000030,
160 corresponding to the Intact_1 prophage region, are actually terminal repeats indicating that the
161 sequence is in fact circular. To confirm this circularity, primers were designed pointing outward
162 from each end of the phage contig (Figure 2A), and PCR and Sanger sequencing was performed
163 on DNA extracted from phage lysates. The sequence of the PCR product obtained was consistent
164 with the conclusion that the phage exists in a circular form at some point during its lifecycle and
165 confirmed complete sequencing of the entire phage genome (Figure 2A).

166 When the phage genome was analysed using BLASTn (14) with the viral filter (taxid:
167 10239), the top hit was Enterobacteria phage Sf101 (accession NC_027398.1) with 96.47%
168 identity but only 33% query cover, indicating that this phage represents a novel viral species, with
169 less than 95% nucleotide sequence similarity to any part of any other characterized phage (15).
170 We thus named this novel phage vB_EcoP_Kap1 (Kap1, NCBI:txid2746235). When the viral
171 filter is removed, the top hit still only shows 96.92% identity and 44% query cover to Kap1
172 (*Escherichia coli* genome assembly FHI87, scaffold-10_contig-14.0_1_42036, accession
173 LM996987.1). Visualization of the alignments between Kap1 and these top BLASTn (14) hits
174 (Figure 2B) showed that a ~10 kb region of Kap1 corresponding to the virion morphogenesis
175 module is the most conserved region. Further, Kap1 may represent a novel genus of the family

176 *Podoviridae*, since it shares less than 50% nucleotide sequence similarity to any other
177 characterized *Podoviridae* genus (15). Comparing the genome sequence of Kapi1 against the type-
178 species for each *Podoviridae* genus in the ICTV 2019.v1 Master Species List (16), the top hit,
179 Enterobacteria phage P22 (accession NC_002371.2), belonging to the genus *Lederbergvirus*,
180 shares only 83.62% identity and 20% query coverage with Kapi1. The taxonomy of Kapi1 was
181 further explored using vContact2 (17), this analysis showed that Kapi1 belongs to the same viral
182 cluster as phages P22 (18) and Sf101 (19) (Figure 2C; Figure S2).

183 The genome of Kapi1 (accession MT813197) is 39,436 bp in length and represents 0.83%
184 of the genome of MP1. The GC content of Kapi1 is 47.1%, slightly lower than the 50.6% of the
185 host genome. Kapi1 has a modular genome structure typical of many lambdoid phages (Figure 2A)
186 (20). Beginning from *xis*, the first region of the Kapi1 genome is rich in hypothetical proteins and
187 proteins with unknown function. The next segment of the genome is characterized by the DNA
188 replication/repair/regulation module; this region has a Lambda-like organization, with CIII, N, CI,
189 *cro*, CII, O, P, and Q. This module is followed by two tRNAs immediately preceding the lysis
190 module (*holin*, *lys*, *Rz*). The next module is responsible for virion morphogenesis, encoding
191 proteins responsible for the head assembly (*terminases*, *scaffold*, *portal*, *coat*), followed by tail
192 assembly (*DNA stabilization protein*, *tail needle knob*, *DNA transfer and ejection proteins*). The
193 final module is required for integration, including *xis*, *int*, and *attP*. For a detailed view of
194 annotation and functional assignments for all CDS in Kapi1, see Table S3.

195

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

197 To begin our search for the integration site of Kapi1, we used BLASTn (14) to look for
198 prophages similar to Kapi1 in the NCBI nucleotide database. We then analyzed the host
199 chromosome surrounding these Kapi1-like prophages to find any similarities with the MP1
200 chromosome. We identified two contigs in MP1 whose ends shared significant similarity to the
201 regions surrounding Kapi1-like prophages found in the NCBI database. The end of the first
202 contig encodes the *dsdAXC* genes, while the end of the second contig encodes *yfdC*, *m1aA*, and
203 *fadLIJ* genes. Primer pairs were designed to amplify the putative prophage-chromosome
204 junctions; PCR products were then sequenced and aligned with the original sequences to
205 determine the integration site of Kapi1, and its orientation in the host chromosome. Kapi1
206 integrates into the chromosome between genes *yfdC* and *dsdC*, with phage *int* gene closest to the
207 chromosomal *yfdC* locus, and the phage *xis* gene on the opposite end of the prophage closest to
208 the chromosomal *dsdC* gene (Figure 3A). Notably, the way in which the original MP1 sequence
209 (11) was assembled actually captures the chromosome-prophage left_junction (Figure 3A), since
210 the end of contig NZ_JEMI01000016 contains Kapi1 *xis* (annotated as *TorI* in the original
211 sequence) directly downstream of *dsdC*. *yfdC* is a predicted inner membrane protein, belonging
212 to the formate-nitrate transporter (FNT) family and may be involved in resistance to surfactants
213 (21); *dsdC* is a transcriptional regulator involved in D-serine detoxification (22). This region of
214 the genome is hypervariable among different *E. coli* pathotypes, and a variety of prophage and
215 phage-like genes are often found at this locus (23).

216 With the integration site for Kapi1 identified, we then looked back at our whole-genome
217 sequencing data for KP7, the strain lacking the Kapi1 prophage. In this strain, the integration locus
218 was correctly assembled; the two contigs that we confirmed to surround the integrated Kapi1
219 prophage in MP1 were assembled into one complete contig in KP7. We noticed a tRNA-Arg in

220 between *dsdC* and *yfdC* that was not annotated on the contigs surrounding the Kapi1 prophage in
221 MP1. Since tRNAs are common integration sites for phage (24), this site was further investigated.
222 When we investigated the chromosomes of MP1 and MP13 with the Kapi1 prophage integrated as
223 described above, we noticed a 17 bp duplication on either end of the integrated prophage; this
224 sequence (5' – AATGGTGTCCCCTGCAG – 3') is found at the 3' end of the tRNA-Arg and is
225 the putative Kapi1 *att* site (Figure 3B). To be clear, this tRNA-Arg is intact whether or not Kapi1
226 is integrated into the chromosome, since the 3' end is maintained by the Kapi1 putative *attP* when
227 it integrates into the chromosome. The *attB* site was not picked up by the auto-annotation programs
228 in MP1 and MP13 since during sequencing the two contigs surrounding Kapi1 were not assembled
229 into the correct scaffold, as they were in KP7.

230 Interestingly, the Kapi1 putative *attP* is identical to prophages Sf6 (25), HK620 (20), and
231 KpIE1 (26) except for the 5' A which is excluded from the Sf6, HK620, and KpIE1 *attP* sites. Like
232 these phages, the Kapi1 *attP* lies between the *int* and *xis* genes, so upon integration into the host
233 chromosome, the *int* and *xis* genes are located on either end of the prophage (Figure 3A). We
234 generated a KP7 *argW::kan* mutant, in this mutant the *attB* core sequence is still intact, but the
235 insertion of the kanamycin resistance cassette disrupts the sequence adjacent to the *attB* core,
236 which may be important for recognition. As expected, this mutant can be infected by Kapi1,
237 although at a reduced frequency when compared to KP7, and produces smaller plaques with an
238 average diameter of 0.89 ± 0.25 mm (Figure 3C). Further experiments showed that KP7 *argW::kan*
239 is still able to be lysogenized by Kapi1, but at a much lower frequency (13.5%) than the wildtype
240 KP7 (100%). Notably, KP7 *argW::kan* lysogens retain the same level of resistance to kanamycin
241 as their non-lysogenic counterparts. The observed reduced rate of lysogeny could be due to a
242 reduced frequency of recognition of the intact *attB* core due to the adjacent sequence disruption

243 by the kanamycin cassette, or possibly integration into other chromosomal sites (27); precise
244 deletion of the putative *attB* core and further experimentation will help to clarify these results.

245

246 The lifestyle of Kapi1 is temperate and shows unstable lysogeny

247 To investigate the lifestyle of Kapi1, we performed phage liberation assays. After only
248 24 hours of growth, $\sim 1 \times 10^8$ PFU/mL can be isolated from standard laboratory cultures of lysogens
249 (data not shown). The ratio of phage per cell (PFU/CFU) is only 0.030 ± 0.010 after 24 hours of
250 growth and rises to 25.80 ± 7.23 PFU/CFU when identical cultures are grown with sub-inhibitory
251 concentrations of mitomycin C (Figure 4A). The mitomycin C assay indicates that Kapi1 can be
252 induced through the traditional SOS pathway in response to DNA damaging agents, although there
253 appears to be a basal level of spontaneous induction in the absence of mitomycin C.

254 In order to study the impact of Kapi1 carriage on growth rate, we compared growth of a
255 Kapi1 lysogen, MP13, to that of the non-lysogenized KP7. Kapi1 slightly reduces the growth rate
256 of the host, and results in a lower cell density in stationary phase (Figure 4B). The cell density of
257 lysogens in stationary phase is also more variable than their non-lysogenic counterparts (see error
258 bars in Figure 4B), likely due to spontaneous induction in a subset of the lysogen population
259 (Figure 4A). The time-course of prophage liberation also agrees with the variation in the stationary
260 phase cell density of lysogens, since Kapi1 seems to be induced at the highest levels during
261 stationary phase (data not shown). This is expected, as stationary phase includes stressful cellular
262 conditions such as nutrient limitation and waste product build-up, resulting in the induction of
263 cellular stress responses, stimulating phage release (for a review of stationary phase in Gram-
264 negative bacteria, see (28)).

265 Since MP1 was recently isolated from the feces of a healthy mouse (11) and is more host-
266 adapted than our standard laboratory strains of *E. coli* such as MG1655 or MC4100, we wondered
267 if Kapi1 might be important to the biology of commensal *E. coli* in the gastrointestinal tract. To
268 investigate the biology of Kapi1 under more physiologically-relevant conditions, we repeated the
269 same prophage liberation experiments in media composed of 50% LB and 50% simulated intestinal
270 fluid (SIF) as well as 50% LB and 50% simulated gastric fluid (SGF) (29). We found lower levels
271 of Kapi1 liberation in SIF when compared to LB (Figure 4A), although the strains grow to nearly
272 identical cell densities. We observed nearly identical levels of Kapi1 liberation in SGF compared
273 to LB (Figure 4A). Prophage stability assays show that upon repeated sub-culturing of a lysogen,
274 the percentage of the population carrying Kapi1 reduces by approximately 10% upon each
275 successive sub-culture in LB, and that the lysogen population is more stable in media composed
276 of 50% LB and 50% SIF (Figure 4C). This result supports our previous finding that the original
277 MP7 isolate lost the Kapi1 prophage before arriving at our lab, likely during handling or passaging.

278 The prophage liberation and stability assays agree; there is a lower level of Kapi1
279 induction when lysogens are grown in LB supplemented with SIF compared to LB alone or LB
280 supplemented with SGF, and this low level of induction is likely responsible for the increased
281 stability of the lysogen population in media supplemented with SIF. This result is unexpected
282 because LB is considered a non-stressful standard lab media, yet there is a higher proportion of
283 spontaneous phage induction in this condition, compared to a less rich (and perhaps more
284 challenging) simulated intestinal media. As well, since the liberation rates in SGF were nearly
285 identical to LB, it seems likely that the lower rate of liberation observed in SIF is not simply due
286 to dilution of the rich LB media and may be specific to intestinal conditions. This could indicate
287 that lysogeny with Kapi1 is selected for in intestinal conditions because it provides some advantage

288 to the cell, which may be dispensable under standard lab conditions. Prophages have been shown
289 to provide fitness advantages to their hosts, including resistance to osmotic, oxidative and acid
290 stresses, as well as influencing biofilm formation (30); many of these stresses would be
291 encountered during colonization of the mammalian gastrointestinal tract. Our observations
292 regarding the stability of Kapi1 lysogens in simulated intestinal conditions warrants further
293 investigation into how (or if) bacteriophage influence *E. coli* colonization of the gastrointestinal
294 tract.

295

296 The Kapi1 receptor is lipopolysaccharide O-antigen

297 Many phage use lipopolysaccharide (LPS) as a receptor for host cell infection, particularly
298 *Podoviridae* (31), including phages HK620 (32) and P22 (18), among others. Based on this
299 knowledge, we hypothesized that Kapi1 may use LPS as its receptor. We isolated spontaneous
300 mutants resistant to Kapi1 by picking survivor colonies from lawns of KP7 overlaid with Kapi1.
301 Since Kapi1 is temperate, survivor colonies were screened for lysogeny by colony PCR with
302 primers spanning the prophage-chromosome junction, and by growing up survivor colonies
303 overnight and spotting their supernatant onto KP7. Two Kapi1-resistant survivor colonies were
304 verified to not be lysogenized by Kapi1, and were chosen for further analysis (KP61, KP62). LPS
305 profiling by SDS-PAGE and silver staining showed that the Kapi1-resistant mutants have severely
306 truncated LPS compared to the WT (Figure 5A). The genomes of KP61 and KP62 were sequenced,
307 and variant calling was performed using WT KP7 as a reference. Both Kapi1-resistant strains
308 contained only one mutation relative to WT; a 1 bp deletion in the *wzy* polymerase causing a
309 frameshift resulting in a truncation (Figure 5B; Figure S3). Mutations in *wzy* result in synthesis of
310 a complete core LPS, but only 1 O-unit is displayed on the cell surface (33), instead of the usual

311 long-chain O-antigens, consistent with what was observed on silver staining. These results indicate
312 that Kapi1 likely binds the KP7 O-antigen as its primary receptor, however it appears that 1 O-unit
313 is insufficient for recognition, as *wzy* mutants are completely resistant to infection (Figure 5C).

314 To confirm that O-antigen is the receptor for Kapi1, we proceeded to create two LPS
315 mutants; KP7 *waaL::kan* which has complete core, but no O-antigen (34), and KP7 *waaF::kan*
316 which has a severely truncated core structure consisting of lipid A, Kdo₂, and heptose (35).
317 Surprisingly, KP7 *waaL::kan* is able to be infected by Kapi1, but at an extremely low efficiency
318 (faint clearing in undiluted spot, Figure 5C), whereas KP7 *waaF::kan* is completely resistant to
319 infection by Kapi1 (Figure 5C). This indicates that in the absence of O-antigen, Kapi1 may be able
320 to recognize a portion of the outer-core structure that is intact in *waaL::kan* but absent in
321 *waaF::kan*, as a secondary receptor; in KP61 and KP62 the single O-unit may obscure this
322 secondary receptor to prevent Kapi1 recognition/binding. Interestingly, *E. coli* W3110 *waaF*
323 mutants have been shown to not produce flagella (36), so this presents another putative secondary
324 receptor for Kapi1. However, no individual plaques can be observed when Kapi1 is spotted on
325 KP7 *waaL::kan* (Figure 5C) so it is also possible that the spot is the result of bactericidal activity,
326 and not a productive phage infection (37). In light of the identification of the O-antigen as the
327 receptor for Kapi1, our previous host-range results make sense; many of the strains we tested lack
328 O-antigen (MC4100, MG1655, BW25113, W3110), and those that do produce O-antigen
329 (E2348/69, J96, DBS100, TJ-LM, TJ-WM, TJ-LR, Nissile 1917) do not appear to have the same
330 O-antigen structure as KP7 on silver-stained SDS-PAGE (data not shown). Although the data
331 represented here is not an extensive screen of all possible serotypes of *E. coli*, it is possible that
332 Kapi1 is specific to one or a few serotypes; identification of the precise region of O-antigen that
333 Kapi1 recognizes as its receptor will help to clarify these results.

334

335 Kap1 modifies the LPS O-antigen

336 While performing LPS profiling in the previous experiments, we noticed that the LPS
337 profiles of MP1 and MP13 differed from the LPS profile of KP7 (Figure 5A). We hypothesized
338 that the change in LPS structure was due to lysogenic conversion by Kap1. Several of the
339 bacteriophage most similar to Kap1 have been shown to cause seroconversion in their respective
340 hosts. Phages Sf101 and Sf6 both encode O-acetyl transferases to cause seroconversion in their
341 host, *Shigella flexneri* (19, 38); and phage P22 encodes an O-antigen glucosylation cassette
342 (*gtrABC*) to cause seroconversion in its host *Salmonella enterica* serovar Typhimurium (39). We
343 began by searching for CDS in the genome of Kap1 with homology to known seroconversion
344 proteins. Although none of the CDS in Kap1 were annotated as possible O-antigen
345 modifying/seroconverting proteins, closer examination revealed limited regions of homology to
346 known seroconversion proteins in a few CDS. Kap1 hypothetical protein 5 (*hyp5*) has limited
347 homology with acyl- and acetyl- transferases from uncultured Mediterranean phage uvMED (57%
348 identity with 68% query coverage, putative acyltransferase [accession BAR19615.1]; 38% identity
349 with 67% query coverage, putative acetyltransferase [accession BAQ85473.1]). Many of the
350 phages infecting *Shigella flexneri* that are closely related to Kap1 encode seroconversion genes
351 that are located near the *int-att-xis* region of the genome (9), so we included two additional CDS
352 as putative seroconversion proteins due to their location in the genome and lack of homology to
353 any known proteins in the NCBI viral BLASTp database. Kap1 hypothetical protein 24 (*hyp24*)
354 is encoded between the tail spike protein and the integrase. The only hit in the NCBI viral database
355 is a hypothetical protein, but when the viral filter (taxid: 10239) is removed, this gene shows
356 significant homology with an acyltransferase from *E. coli* (79% identity with 94% query coverage,

357 TPA: acyltransferase [*Escherichia coli*], accession HAH9668903.1). Kapi1 hypothetical protein
358 25 (*hyp25*) is encoded between the integrase and *attP* site; the only hits in the NCBI database (with
359 or without viral filter) are hypothetical proteins.

360 We cloned each of these three putative seroconversion proteins into the pTrc99a
361 overexpression vector (40), and introduced them into KP7, then compared the LPS profiles of each
362 overexpression strain, the vector control, and WT KP7 and MP13. Unfortunately, none of the
363 strains overexpressing the three putative seroconversion proteins from Kapi1 had altered LPS
364 profiles (data not shown). Although none of these proteins were found to be individually
365 responsible for seroconversion, we cannot eliminate the possibility that they could be working
366 with other phage-encoded proteins to produce the altered LPS phenotype. There are numerous
367 examples of seroconverting phage that encode entire gene cassettes responsible for
368 seroconversion, including *P. aeruginosa* phage D3 (41) and *Salmonella* phage P22 (39). An
369 alternative approach would be to individually delete putative seroconversion proteins from Kapi1
370 and assay lysogens for O-antigen structural changes back to the WT form. It is interesting,
371 however, that the genetic basis for O-antigen modification by Kapi1 appears to not be well-
372 conserved, as no strong hits to known O-antigen modifying proteins could be identified. It will be
373 valuable to determine the molecular mechanisms behind Kapi1 phage-mediated O-antigen
374 modification, and whether these mechanisms are indeed novel.

375 Phages that use the O-antigen as a receptor also commonly modify the O-antigen using the
376 tailspike protein (*tsp*) to facilitate movement to the bacterial outer membrane, where irreversible
377 binding and particle opening can occur (42). To determine if O-antigen degradation via *tsp* was
378 responsible for the altered LPS structure between Kapi1 lysogens and non-lysogens observed on
379 SDS-PAGE (Figure 5A), we set up a mock infection with Kapi1 and purified KP7 LPS and ran

380 these samples alongside the uninfected controls from KP7 and MP13. It appears that treatment of
381 purified KP7 LPS with Kapi1 results in an altered LPS structure (suggesting degradation of O-
382 antigen by Kapi1), but this structure is different from that of either the WT or lysogenic
383 backgrounds (Figure 5D). Therefore, Kapi1 is responsible for alteration of the LPS structure both
384 upon binding O-antigen prior to infection, and later via lysogenic conversion. Kapi1 *tsp* has
385 considerable sequence conservation in the head-binding domain (most similar to phages HK620
386 and Sf101), but no sequence similarity to any other viral proteins in the NCBI database along the
387 length of the protein. Therefore, it is difficult to predict what type of enzymatic activity the Kapi1
388 *tsp* may have, and further work is needed to characterize the molecular mechanisms of Kapi1-
389 mediated O-antigen modification through *tsp*.

390 In conclusion, we have isolated and begun the characterization of a novel bacteriophage
391 infecting commensal *E. coli*. The genome of Kapi1 has been sequenced and annotated, and the
392 prophage integration site in the host genome has been identified. Further, we demonstrate that
393 Kapi1 shows unstable lysogeny, and lysogeny appears to be selected for in intestinal conditions.
394 O-antigen is the Kapi1 receptor and Kapi1 appears to modify the host O-antigen upon initial
395 binding, and later in infection through lysogenic conversion, although the molecular mechanisms
396 are yet to be elucidated. Our findings suggest that Kapi1 lysogeny may confer some advantage
397 during colonization of the intestine and demonstrate that this novel temperate phage alters the
398 structure of LPS, a major determinant of bacterial interaction with the immune system. MP1 and
399 Kapi1 will serve as a good model system to explore what role(s) temperate phages may play in
400 colonization of the gastrointestinal tract by commensal strains of *E. coli*.

401

402 **Materials and Methods**

403 Bacterial Strains and Growth Conditions

404 Strains MP1, MP7, and MP13 were a generous gift from the Goulian lab (11). A complete
405 list of bacterial strains used in this study can be found in Table S1. Unless otherwise specified, all
406 strains were grown in LB (10 g/L tryptone, 5 g/L NaCl, 5 g/L yeast extract) supplemented with
407 the appropriate antibiotics, at 37 °C and 220 rpm. When plated, cells were grown on LB 1.5% agar
408 supplemented with the appropriate antibiotics and inverted at 37 °C. Antibiotic concentrations
409 used are as follows: ampicillin 100 µg/mL, kanamycin 30 or 50 µg/mL.

410

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

412 Phage were isolated from overnight cultures of MP1 or MP13 by pelleting cells, and filter-
413 sterilizing the supernatant using a 0.45 µm syringe-driven filter. Individual plaques were isolated,
414 propagated, and phage stocks prepared by previously described methods (43), slightly modified.
415 Briefly, the above phage-containing supernatant was mixed with susceptible host strain KP7 1:1,
416 then 3.5 mL soft agar (LB 0.7% agar) was added, the mixture was poured onto solid LB agar
417 plates, and incubated overnight. Individual plaques were picked with a sterile Pasteur pipette and
418 gently resuspended in 500 µL modified suspension media (SM) (50 mM Tris-HCl (pH 7.4),
419 100 mM NaCl, 10 mM MgSO₄), this suspension was then amplified using the soft agar overlay
420 technique as above. Plates with near-confluent lysis were used to prepare high-titre stocks by
421 collecting the soft agar layer as follows: SM was poured onto the surface of the plate, and soft agar
422 was gently scraped into a 50 mL Falcon tube using a sterile scoopula, rocked at room temperature
423 for 1 hour and centrifuged to pellet the soft agar. The supernatant was filter-sterilized using a 0.22

424 μm syringe-driven filter, and stored at 4 °C. Plaque diameter was measured from 10 plaques, using
425 Fiji software (44); the mean and standard deviation are reported.

426 Phage samples were prepared for transmission electron microscopy (TEM) and imaged by
427 previously described methods using uranyl acetate as a background stain (43) at the University of
428 Alberta Advanced Microscopy Facility. Virion measurements were performed using Fiji software
429 (44) from 44 phage particles; the mean and standard deviation are reported.

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

439

440 Genome Sequencing and PCR

441 Pure cultures were grown by streaking from glycerol cryo-stocks, then picking single
442 colonies and growing in LB broth overnight. DNA was extracted from overnight cultures using
443 the Lucigen MasterPure Complete DNA and RNA Purification Kit, and the concentration and
444 quality of gDNA was checked using the NanoDrop 2000c. Library preparation and whole genome
445 sequencing was performed by the Microbial Genome Sequencing Centre (MiGS, Pittsburgh PA).

446 Libraries were prepared with the Illumina Nextera kit, and sequenced using the NextSeq 550
447 platform.

448 Standard PCR was performed using Taq polymerase (Invitrogen) following the
449 manufacturer's directions; a single colony was suspended in 20 μ L nuclease-free water and 5 μ L
450 of this suspension was used per 50 μ L reaction. Phage DNA was extracted using a standard
451 phenol/chloroform extraction protocol from the Center for Phage Technology at Texas A&M (45),
452 and 1-5 μ L of phage DNA was used per 50 μ L reaction. PCR products were checked by running
453 10 μ L on a 2% agarose gel and staining with ethidium bromide. All PCR primers used in this study
454 can be found in Table S2. Sanger sequencing was performed by the Molecular Biology Service
455 Unit at the University of Alberta.

456

457 Genome Assembly, Annotation, and Taxonomy

458 Paired-end reads of 2x150 bp received from MiGS were uploaded to the public server at
459 usegalaxy.org (46). Galaxy, and all tools there-in, was used for bacterial genome assembly,
460 annotation, and analysis as follows. Illumina adapters and low-quality reads were trimmed using
461 Trim Galore! (47), reads were then checked for quality using FastQC (48) and MultiQC (49).
462 Trimmed reads were *de novo* assembled using Unicycler (50), functioning as a SPAdes (51)
463 optimizer as no long-read data was generated. Quality of assemblies was assessed using Quast
464 (52), and bacterial genomes were then annotated using Prokka (53, 54). SnapGene software (from
465 Insightful Science; available at snapgene.com) and Geneious Prime 2019.2.3
466 (<http://www.geneious.com/>) were used for genome visualization, creating genome maps, and
467 designing primers.

468 The Kapi1 genome was manually annotated using previously described methods (55),
469 slightly modified. This method uses a rigorous approach to score potential coding sequences
470 (CDS) based on various parameters; low-scoring CDS are then discarded, and the remaining CDS
471 are analyzed to determine their correct start codon, again based on a scoring system. Briefly, the
472 prophage genome was run through three auto-annotation programs; GenemarkS (56), Glimmer3
473 (57) at CPT Phage Galaxy public server (cpt.tamu.edu/galaxy-pub), and Prokka (53) at
474 usegalaxy.org (54). The coding potential for each putative CDS was determined using GenemarkS
475 coding potential graph (56). Putative CDS were then searched against NCBI's non-redundant
476 protein database using BLASTp (14), and scored based on whether they had significantly similar
477 hits (as determined by the E-value), and whether those hits were known proteins or hypotheticals.
478 Each CDS was scored based on the length of overlap with neighboring CDS, as extremely long
479 overlaps are unlikely, while short overlaps of 1, 4 or 8 bp are more favorable as these suggest
480 organization into an operon. Finally, CDS were scored based on the length of the ORF, where
481 extremely short CDS are penalized the most heavily. Low scoring CDS were discarded prior to
482 start codon identification. Start codons for each CDS were scored in a similar manner, using again
483 the coding potential graph from GenemarkS (56), the number of auto-annotation programs that
484 selected the start codon, sequence similarity matches in NCBI, and length of the ORF. These
485 parameters are listed in order of most important to least and were used to select the most likely
486 start codon for each CDS. CD-Search (58) was also performed for all CDS to assist with functional
487 assignment.

488 Taxonomic evaluation was performed using vContact2 v9.8 (17) through the CyVerse
489 platform (www.cyverse.org). The analysis was run with the default parameters, using NCBI
490 Bacterial and Archaeal Viral RefSeq V85 (with ICTV + NCBI taxonomy) as the reference

491 database. The resulting network was visualized using CytoScape (59). Duplicated edges were
492 removed from the network (edges represent connections between two nodes, in this case viruses),
493 and only first-neighbors to Kapi1 were kept (nodes that have a direct connection to the Kapi1
494 node). An edge-weighted, spring-embedded layout was used so that nodes that are more closely
495 related appear closer together spatially, and edges were weighted so that stronger connections (ie.
496 more sequence similarity between two viruses) appear darker and thicker.

497

498 Rates of phage liberation and prophage stability assays

499 The rate of phage liberation was determined by enumerating CFUs and PFUs in cultures
500 of MP13 Kapi1 lysogens over time. Three colonies of MP13 and three colonies of KP7 (non-
501 lysogenic control) were picked and grown overnight at 37 °C, the next day cultures were adjusted
502 to OD₆₀₀ 1.0 to ensure equal cell numbers, then sub-cultured 1:100 into LB, LB with 0.5 ng/μL
503 mitomycin C, LB mixed 50-50 with simulated intestinal fluid (SIF – 6.8 g KH₂PO₄, 1.25 g
504 pancreatin, 3 g bile salts in 1 L dH₂O, pH adjusted to 7 (29)), or LB mixed 50-50 with simulated
505 gastric fluid (SGF – 2 g/L NaCl, 3.2 g/L porcine mucosa pepsin, pH adjusted to 3.5 (29)). After
506 24 hours incubation, an aliquot was taken from each culture, cells were spun down, washed in
507 phosphate-buffered saline (PBS) (137 mM NaCl, 2.7 mM KCl, 10 mM Na₂HPO₄, 1.8 mM
508 KH₂PO₄), serially diluted, plated onto LB, and grown overnight to enumerate the number of cells
509 in the culture. In parallel, the culture supernatants (containing phage) were serially diluted, spotted
510 onto KP121 soft agar overlays, and incubated overnight at 30 °C to enumerate the number of phage
511 particles in the culture. Phage released per cell was estimated by dividing PFU/mL by CFU/mL.
512 KP121 was used to enumerate phage as this strain has a much lower efficiency of lysogeny with
513 Kapi1 compared to WT KP7, allowing for more accurate enumeration.

514 Efficiency of lysogeny was determined by infecting KP7 and KP121 with Kapi1 at a
515 multiplicity of infection of 10 at 37 °C for 30 minutes, then washing and plating out surviving
516 cells. The following day, 30 survivor colonies were picked and grown up in liquid culture
517 overnight. Overnight cultures were assayed for lysogeny with Kapi1 in two ways: first, soft-agar
518 overlays were prepared with each survivor culture, and spotted with Kapi1; second, survivor cells
519 were spun down, and the supernatant was spotted onto a soft-agar overlay prepared with KP7.
520 Kapi1 lysogens are resistant to lysis by Kapi1, and their supernatants produce plaques on KP7.

521 Prophage stability was assayed by serially propagating cultures of MP13 Kapi1 lysogens.
522 Cultures were grown in biological triplicates (three independent colonies) in either LB, LB mixed
523 50-50 with SGF, LB mixed 50-50 with SIF, or LB with 0.5 ng/μL mitomycin C for 24 hours, then
524 sub-cultured into fresh media and grown for another 24 hours. After a total of 48 hours incubation
525 (1 passage) cells were spun down and washed twice in PBS, serially diluted, plated on LB and
526 grown overnight to get individual colonies. Using velvet squares, colonies were replicated onto a
527 LB plate spread with 50 μL of an overnight culture of KP7 to screen for lysogens (60). This replica
528 plating technique results in two distinct phenotypes: colonies that produce a zone of clearing in
529 the KP7 lawn are scored as lysogens, and colonies without a zone of clearing are scored as non-
530 lysogenic (see Figure S4 for examples and experimental verification of these phenotypes). The
531 relative loss of Kapi1 lysogens in the culture was calculated by dividing the non-lysogen CFU/mL
532 by the total CFU/mL.

533

534 Isolation of Spontaneous Kapi1-resistant mutants

535 Spontaneous Kapi1-resistant mutants were isolated by spotting Kapi1 onto a soft-agar
536 overlay prepared with KP7. The following day, six colonies that grew within the cleared phage
537 spots were picked and re-streaked onto LB. The following day, these colonies were screened for
538 lysogeny with Kapi1 using PCR with primers that span the phage-genome junction. Two colonies
539 that did not produce phage-genome bands were grown up overnight, then the supernatant was
540 filter-sterilized and spotted onto a soft-agar overlay prepared with KP7. The absence of plaques
541 on KP7 confirmed that these two mutants are not Kapi1 lysogens. To identify which mutations
542 were responsible for the Kapi1-resistant phenotype, each colony was sent for whole genome
543 sequencing at MiGS, as above. SNPs were identified using BreSeq (61) and snippy (62) through
544 Galaxy (46), using the KP7 genome as a reference.

545

546 Generation of Mutants

547 To create *waaF* mutants, P1 transduction was performed as previously described (63, 64)
548 using the corresponding Keio collection mutant (65) as a donor. The Lambda Red system (66) was
549 used to generate KP7 *waaL::kan*, and KP7 *argW::kan*. Briefly, primers were designed using 50-
550 ntd 5' extensions homologous to the regions surrounding the gene of interest (H1/H2) with 20-ntd
551 3' extensions homologous to the regions flanking the FRT sites and Kan cassette in pKD13 (P1/P2)
552 (65). PCR was performed with Phusion polymerase to amplify the FRT-flanked Kan cassette from
553 pKD13, then these products were purified using the QIAquick PCR Purification Kit and
554 electroporated into KP7 carrying the helper plasmid pKD46. Recovery was performed at 37 °C for
555 2-3 hours, cells were then plated on LB with 30 µg/mL kanamycin and grown at 37 °C for 24
556 hours to select for recombinant strains. Kanamycin-resistant colonies were then struck onto LB
557 with 50 µg/mL kanamycin and grown overnight at 37 °C, then patched onto LB with or without

558 ampicillin to screen for loss of pKD46. Kanamycin-resistant and ampicillin-sensitive strains were
559 confirmed by colony PCR using primers designed to span the upstream and downstream junctions
560 between the genome and kanamycin cassette, and primers that spanned the entire disrupted region.
561 PCR products were sequenced at the University of Alberta Molecular Biology Service Unit to
562 confirm the desired mutations. Once sequencing confirmed the correct mutation, P1 transduction
563 was used (as above) to move these mutations into a fresh KP7 background, to avoid the possibility
564 of any off-site mutations acquired during construction. In all cases, strains were confirmed to be
565 non-lysogenic for Kapi1 by PCR and spotting assays, as above.

566

567 Lipopolysaccharide Profiling

568 The general structure of lipopolysaccharide (LPS) was analyzed by profiling LPS extracts
569 on SDS-PAGE with silver staining. LPS was extracted using a modified proteinase K micro-
570 digestion protocol (67) as follows; bacterial strains of interest were grown overnight, then 1 mL
571 of this culture was washed twice with PBS, resuspended to a final OD₆₀₀ of 2.0 in PBS, and
572 pelleted. Cells were resuspended in 50 µL lysis buffer (2% SDS, 4% 2-mercaptoethanol, 10%
573 glycerol, 1M Tris pH 6.8, bromophenol blue to a deep blue color) and incubated at 95 °C for 10
574 minutes to lyse cells. Whole cell lysate was cooled to room temperature, then 10 µL 2.5 mg/mL
575 proteinase K (20 mg/mL stock solution was diluted in lysis buffer first) was added and incubated
576 at 56 °C for 1 hour. Standard polyacrylamide gels were prepared with 12% acrylamide (19:1
577 acrylamide:bisacrylamide) (68), 1-5 µL of proteinase K-treated whole cell lysate (LPS extract)
578 was loaded per well and run at 20 mA constant current in Tris-glycine running buffer (25 mM Tris,
579 200 mM glycine, 0.1% SDS) until the dye front nearly reached the bottom of the gel. Silver staining

580 was performed as previously described (69), and imaged on a clear petri dish using an iPhone
581 camera.

582

583 Cloning

584 Putative seroconversion proteins were cloned into pTrc99a (40) using a basic PCR-based
585 cloning method. Coding sequences for the candidate seroconversion proteins were amplified from
586 MP13 colonies using Phusion polymerase; forward primers were designed with an EcoRI
587 restriction site, and reverse primers included a BamHI restriction site. Four 50 μ L PCR reactions
588 were pooled for each gene, then cleaned up using the QIAquick PCR Purification Kit. Purified
589 PCR products and empty pTrc99a were double digested with EcoRI and BamHI FastDigest
590 enzymes (Thermo Scientific). Digested vector and inserts were again pooled and cleaned up using
591 the QIAquick PCR Purification Kit. Inserts were ligated into pTrc99a using a 3:1 molar ratio, using
592 T4 DNA ligase (Invitrogen). 8 μ L of each ligation was transformed into One Shot TOP10
593 Chemically Competent *E. coli* (Invitrogen) and were selected on LB plates containing ampicillin.
594 Ampicillin-resistant colonies were screened for inserts by PCR, using primers binding just outside
595 of the multiple cloning site in pTrc99a; colonies with the correct size insert were grown up
596 overnight, the plasmids were isolated using QIAprep Spin Miniprep Kit, then sequenced at the
597 University of Alberta Molecular Biology Service Unit. After verification of the sequences,
598 plasmids were electroporated into KP7. Briefly, electrocompetent cells were prepared by
599 harvesting mid-log phase cells, washing four times in ice-cold 10% glycerol and concentrating
600 cells 500-fold. Electrocompetent cells were combined with 1-5 μ L plasmid DNA, electroporated
601 at 2.5 kV in a 0.2 mm cuvette, immediately resuspended in 1 mL LB, recovered for 2-3 hours at
602 37 °C with shaking, and then plated on LB with ampicillin to select for transformants. To screen

603 candidate seroconversion proteins, KP7 transformants carrying pTrc99a constructs were sub-
604 cultured 1:50 into LB with ampicillin and grown to an OD₆₀₀ of ≥ 0.5 , then IPTG was added to a
605 final concentration of 1 mM and cells were grown until an OD₆₀₀ of ≥ 1.0 (20-60 mins of induction)
606 before LPS extraction, SDS-PAGE, and silver staining, as above. All transformants were verified
607 to be non-lysogenic for Kapi1 using the replica plating method described above.

608

609 Lipopolysaccharide Degradation

610 To determine if Kapi1 is capable of degrading LPS, LPS was extracted from cultures of
611 KP7 (non-lysogenic for Kapi1) and MP13 Kapi1 lysogens using a phenol extraction method
612 adapted from Davis and Goldberg (70), purified KP7 LPS was then incubated with Kapi1 and
613 compared to the untreated controls. The phenol extraction was used in place of proteinase K
614 digestion (as above), as Kapi1 was not viable in the lysis buffer, even after inactivation of
615 proteinase K at high temperatures (data not shown). Briefly, KP7 and MP13 were grown up
616 overnight, cultures were pelleted and washed in PBS, then resuspended in 1.5 mL PBS to give a
617 final OD⁶⁰⁰ of 2.0. Cells were pelleted and resuspended in 200 μ L Laemmli buffer (50 mM Tris-
618 HCl pH 6.8, 4% SDS, 10% glycerol, 0.1% bromophenol blue, 5% β -mercaptoethanol), then boiled
619 for 15 minutes to lyse cells. Once cool, DNase I and RNase A were added to cell lysates at 37 °C
620 for 10 minutes. Proteinase K was then added, and lysates were incubated at 55 °C overnight. The
621 following day, 200 μ L Tris-saturated phenol was added to lysates, then vortexed for 10 seconds
622 before incubating at 65 °C for 15 minutes. Once cool, samples were centrifuged at 14,000 rpm for
623 10 min at 4 °C, and the upper phase was transferred to a new tube. 2.5 vol ethanol was added to
624 precipitate LPS, then centrifuged at 15,000 rpm for 20 minutes. Supernatant was discarded, pellets
625 air dried, then resuspended in 50 μ L nuclease-free water. Kapi1 was added to KP7 LPS at a “MOI”

626 of 10 (assuming that 1.5 mL of OD⁶⁰⁰ 2.0 culture was concentrated into 50 µL) and incubated at
627 37 °C without shaking for 30 minutes. Phage-treated and untreated KP7 LPS, along with MP13
628 untreated LPS were run on SDS-PAGE and silver stained, as above, using 15 µL of LPS extract,
629 as this extraction method produced a lower yield. A portion of the phage-treated samples were
630 serially diluted and spotted onto an KP7 soft-agar overlay to ensure that phage particles remained
631 viable after incubation with LPS (data not shown).

632

633 Data Availability

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

635

636 **Acknowledgments**

637 The authors thank Arlene Oatway from the University of Alberta Advanced Microscopy
638 Facility for assistance with transmission electron microscopy; the University of Alberta
639 Molecular Biology Facility for assistance with Sanger sequencing; the Microbial Genome
640 Sequencing Centre for assistance with whole genome sequencing; Dr. Mark Goulian for critical
641 reading of the manuscript and for strains MP1, MP7, MP13; Dr. Ben Willing for strains TJ-LM,
642 TJ-WM, and TJ-LR; Jaclyn McCutcheon for advice on phage isolation and preliminary
643 characterization; and Dr. Brent Weber for advice on identification of the integration site.

644 This research was supported by operating grants from The National Sciences and
645 Engineering Research Council (NSERC) and The Canadian Institutes of Health Research (CIHR),
646 and a project grant from the AMR – One Health Consortium, funded by the Major Innovation Fund

647 program of the Ministry of Jobs, Economy and Innovation, Government of Alberta, to TR. KP was
648 supported by an NSERC Alexander Graham Bell Canada Graduate Scholarship - Master's, Walter
649 H Johns Graduate Fellowship, University of Alberta Science Graduate Scholarship, and Susan
650 Eberlein Graduate Scholarship in Genetics.

651

652 **References**

- 653 1. Tenaillon O, Skurnik D, Picard B, Denamur E. 2010. The population genetics of
654 commensal *Escherichia coli*. *Nat Rev Microbiol* 8:207–217.
- 655 2. Howard-Varona C, Hargreaves KR, Abedon ST, Sullivan MB. 2017. Lysogeny in nature:
656 Mechanisms, impact and ecology of temperate phages. *ISME J* 11:1511–1520.
- 657 3. Touchon M, Bernheim A, Rocha EPC. 2016. Genetic and life-history traits associated
658 with the distribution of prophages in bacteria. *ISME J* 10:2744–2754.
- 659 4. Kim MS, Bae JW. 2018. Lysogeny is prevalent and widely distributed in the murine gut
660 microbiota. *ISME J* 12:1127–1141.
- 661 5. Hsu BB, Gibson TE, Yeliseyev V, Liu Q, Lyon L, Bry L, Silver PA, Gerber GK. 2019.
662 Dynamic Modulation of the Gut Microbiota and Metabolome by Bacteriophages in a
663 Mouse Model. *Cell Host Microbe* 25:803-814.e5.
- 664 6. Frazão N, Sousa A, Lässig M, Gordo I. 2019. Horizontal gene transfer overrides mutation
665 in *Escherichia coli* colonizing the mammalian gut. *Proc Natl Acad Sci U S A* 116:17906–
666 17915.
- 667 7. Davies E V., Winstanley C, Fothergill JL, James CE. 2016. The role of temperate

- 668 bacteriophages in bacterial infection. *FEMS Microbiol Lett.*
- 669 8. Van Belleghem JD, Dąbrowska K, Vaneechoutte M, Barr JJ, Bollyky PL. 2019.
- 670 Interactions between bacteriophage, bacteria, and the mammalian immune system. *Viruses*
- 671 11.
- 672 9. Allison GE, Verma NK. 2000. Serotype-converting bacteriophages and O-antigen
- 673 modification in *Shigella flexneri*. *Trends Microbiol* 8:17–22.
- 674 10. Guest RL, Rutherford ST, Silhavy TJ. 2020. Border Control : Regulating LPS Biogenesis.
- 675 *Trends Microbiol* 1–12.
- 676 11. Lasaro M, Liu Z, Bishar R, Kelly K, Chattopadhyay S, Paul S, Sokurenko E, Zhu J,
- 677 Goulian M. 2014. *Escherichia coli* isolate for studying colonization of the mouse intestine
- 678 and its application to two-component signaling knockouts. *J Bacteriol* 196:1723–1732.
- 679 12. Arndt D, Grant JR, Marcu A, Sajed T, Pon A, Liang Y, Wishart DS. 2016. PHASTER: a
- 680 better, faster version of the PHAST phage search tool. *Nucleic Acids Res* 44:W16–W21.
- 681 13. Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS. 2011. PHAST: A Fast Phage Search
- 682 Tool. *Nucleic Acids Res* 39:347–352.
- 683 14. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment
- 684 search tool. *J Mol Biol* 215:403–410.
- 685 15. Adriaenssens EM, Rodney Brister J. 2017. How to name and classify your phage: An
- 686 informal guide. *Viruses* 9:1–9.
- 687 16. International Committee on Taxonomy of Viruses. 2020. Master Species Lists. Retrieved
- 688 from <https://talk.ictvonline.org/files/master-species-lists/>.

- 689 17. Bin Jang H, Bolduc B, Zablocki O, Kuhn JH, Roux S, Adriaenssens EM, Brister JR,
690 Kropinski AM, Krupovic M, Lavigne R, Turner D, Sullivan MB. 2019. Taxonomic
691 assignment of uncultivated prokaryotic virus genomes is enabled by gene-sharing
692 networks. *Nat Biotechnol* 37:632–639.
- 693 18. Neal BL, Brown PK, Reeves PR. 1993. Use of Salmonella phage P22 for transduction in
694 *Escherichia coli*. *J Bacteriol* 175:7115–7118.
- 695 19. Jakhetia R, Marri A, Ståhle J, Widmalm G, Verma NK. 2014. Serotype-conversion in
696 *Shigella flexneri*: Identification of a novel bacteriophage, Sf101, from a serotype 7a strain.
697 *BMC Genomics* 15.
- 698 20. Clark AJ, Inwood W, Cloutier T, Dhillon TS. 2001. Nucleotide sequence of coliphage
699 HK620 and the evolution of lambdoid phages. *J Mol Biol* 311:657–679.
- 700 21. Nakata K, Koh MM, Tsuchido T, Matsumura Y. 2010. All genomic mutations in the
701 antimicrobial surfactant-resistant mutant, *Escherichia coli* OW66, are involved in cell
702 resistance to surfactant. *Appl Microbiol Biotechnol* 87:1895–1905.
- 703 22. McFall E, Heincz M. 1983. Identification and control of synthesis of the *dsdC* activator
704 protein. *J Bacteriol* 153:872–877.
- 705 23. Moritz RL, Welch RA. 2006. The *Escherichia coli* *argW-dsdCXA* genetic island is highly
706 variable, and *E. coli* K1 strains commonly possess two copies of *dsdCXA*. *J Clin*
707 *Microbiol* 44:4038–4048.
- 708 24. Reiter W-D, Palm P, Yeats S. 1989. Transfer RNA genes frequently serve as integration
709 sites for prokaryotic genetic elements. *Nucleic Acids Res* 17.

- 710 25. Casjens S, Winn-Stapley DA, Gilcrease EB, Morona R, Kühlewein C, Chua JEH,
711 Manning PA, Inwood W, Clark AJ. 2004. The chromosome of *Shigella flexneri*
712 bacteriophage Sf6: Complete nucleotide sequence, genetic mosaicism, and DNA
713 packaging. *J Mol Biol* 339:379–394.
- 714 26. Panis G, Méjean V, Ansaldi M. 2007. Control and regulation of KplE1 prophage site-
715 specific recombination: A new recombination module analyzed. *J Biol Chem* 282:21798–
716 21809.
- 717 27. Barreiro V, Haggard-Ljungquist E. 1992. Attachment sites for bacteriophage P2 on the
718 *Escherichia coli* chromosome: DNA sequences, localization on the physical map, and
719 detection of a P2-like remnant in *E. coli* K-12 derivatives. *J Bacteriol* 174:4086–4093.
- 720 28. Navarro Llorens JM, Tormo A, Martínez-García E. 2010. Stationary phase in gram-
721 negative bacteria. *FEMS Microbiol Rev*. Blackwell Publishing Ltd.
- 722 29. Millette M, Nguyen A, Amine KM, Lacroix M. 2013. Gastrointestinal survival of bacteria
723 in commercial probiotic products. *Int J Probiotics Prebiotics* 8:149–156.
- 724 30. Wang X, Kim Y, Ma Q, Hong SH, Pokusaeva K, Sturino JM, Wood TK. 2010. Cryptic
725 prophages help bacteria cope with adverse environments. *Nat Commun* 1:147.
- 726 31. Silva JB, Storms Z, Sauvageau D. 2016. Host receptors for bacteriophage adsorption.
727 *FEMS Microbiol Lett* 363:1–11.
- 728 32. Dhillon TS, Poon APW, Chan D, Clark AJ. 1998. General transducing phages like
729 *Salmonella* phage P22 isolated using a smooth strain of *Escherichia coli* as host. *FEMS*
730 *Microbiol Lett* 161:129–133.

- 731 33. Samuel G, Reeves P. 2003. Biosynthesis of O-antigens: Genes and pathways involved in
732 nucleotide sugar precursor synthesis and O-antigen assembly. *Carbohydr Res* 338:2503–
733 2519.
- 734 34. Heinrichs DE, Monteiro MA, Perry MB, Whitfield C. 1998. The assembly system for the
735 lipopolysaccharide R2 core-type of *Escherichia coli* is a hybrid of those found in
736 *Escherichia coli* K-12 and *Salmonella enterica*. Structure and function of the R2 WaaK
737 and WaaL homologs. *J Biol Chem* 273:8849–8859.
- 738 35. Gronow S, Brabetz W, Brade H. 2000. Comparative functional characterization in vitro of
739 heptosyltransferase I (WaaC) and II (WaaF) from *Escherichia coli*. *Eur J Biochem*
740 267:6602–6611.
- 741 36. Wang Z, Wang J, Ren G, Li Y, Wang X. 2016. Deletion of the genes *waaC*, *waaF*, or
742 *waaG* in *Escherichia coli* W3110 disables the flagella biosynthesis. *J Basic Microbiol*
743 56:1021–1035.
- 744 37. Abedon ST. 2018. Detection of Bacteriophages: Phage Plaques, p. 1–32. *In*
745 *Bacteriophages*.
- 746 38. Verma NK, Brandt JM, Verma DJ, Lindberg AA. 1991. Molecular characterization of the
747 O-acetyl transferase gene of converting bacteriophage SF6 that adds group antigen 6 to
748 *Shigella flexneri*. *Mol Microbiol* 5:71–75.
- 749 39. Vander Byl C, Kropinski AM. 2000. Sequence of the genome of *Salmonella*
750 bacteriophage P22. *J Bacteriol* 182:6472–6481.
- 751 40. Amann E, Ochs B, Abel KJ. 1988. Tightly regulated *tac* promoter vectors useful for the

- 752 expression of unfused and fused proteins in *Escherichia coli*. *Gene* 69:301–315.
- 753 41. Newton GJ, Daniels C, Burrows LL, Kropinski AM, Clarke AJ, Lam JS. 2001. Three-
754 component-mediated serotype conversion in *Pseudomonas aeruginosa* by bacteriophage
755 D3. *Mol Microbiol* 39:1237–1247.
- 756 42. Broecker NK, Barbirz S. 2017. Not a barrier but a key: How bacteriophages exploit host's
757 O-antigen as an essential receptor to initiate infection. *Mol Microbiol* 105:353–357.
- 758 43. Peters DL, Mccutcheon JG, Stothard P, Dennis JJ. 2019. Novel *Stenotrophomonas*
759 maltophilia temperate phage DLP4 is capable of lysogenic conversion. *BMC Genomics*
760 20.
- 761 44. Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S,
762 Rueden C, Saalfeld S, Schmid B, Tinevez JY, White DJ, Hartenstein V, Eliceiri K,
763 Tomancak P, Cardona A. 2012. Fiji: An open-source platform for biological-image
764 analysis. *Nat Methods* 9:676–682.
- 765 45. Center for Phage Technology. 2018. Protocol for Phage DNA Extraction with
766 Phenol:Chloroform. Retrieved from <https://cpt.tamu.edu/phage-links/phage-protocols>.
- 767 46. Afgan E, Baker D, Batut B, Van Den Beek M, Bouvier D, Ech M, Chilton J, Clements D,
768 Coraor N, Grüning BA, Guerler A, Hillman-Jackson J, Hiltmann S, Jalili V, Rasche H,
769 Soranzo N, Goecks J, Taylor J, Nekrutenko A, Blankenberg D. 2018. The Galaxy
770 platform for accessible, reproducible and collaborative biomedical analyses: 2018 update.
771 *Nucleic Acids Res* 46:W537–W544.
- 772 47. Babraham Bioinformatics. 2012. Trim Galore! A wrapper tool around Cutadapt and

- 773 FastQC to consistently apply quality and adapter trimming to FastQ files. Retrieved from
774 https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/.
- 775 48. Babraham Bioinformatics. 2010. FastQC, A quality control tool for high throughput
776 sequence data. Retrieved from <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.
- 777 49. Ewels P, Magnusson M, Lundin S, Källér M. 2016. MultiQC: Summarize analysis results
778 for multiple tools and samples in a single report. *Bioinformatics* 32:3047–3048.
- 779 50. Wick RR, Judd LM, Gorrie CL, Holt KE. 2017. Unicycler: Resolving bacterial genome
780 assemblies from short and long sequencing reads. *PLoS Comput Biol* 13:1–16.
- 781 51. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM,
782 Nikolenko SI, Pham S, Prjibelski AD, Pyshkin A V., Sirotkin A V., Vyahhi N, Tesler G,
783 Alekseyev MA, Pevzner PA. 2012. SPAdes: A new genome assembly algorithm and its
784 applications to single-cell sequencing. *J Comput Biol* 19:455–477.
- 785 52. Mikheenko A, Prjibelski A, Saveliev V, Antipov D, Gurevich A. 2018. Versatile genome
786 assembly evaluation with QUAST-LG. *Bioinformatics* 34:i142–i150.
- 787 53. Seemann T. 2014. Prokka: Rapid prokaryotic genome annotation. *Bioinformatics*
788 30:2068–2069.
- 789 54. Cuccuru G, Orsini M, Pinna A, Sbardellati A, Soranzo N, Travaglione A, Uva P, Zanetti
790 G, Fotia G. 2014. Orione, a web-based framework for NGS analysis in microbiology.
791 *Bioinformatics* 30:1928–1929.
- 792 55. Salisbury A, Tsourkas PK. 2019. A method for improving the accuracy and efficiency of
793 bacteriophage genome annotation. *Int J Mol Sci* 20.

- 794 56. Besemer J. 2001. GeneMarkS: a self-training method for prediction of gene starts in
795 microbial genomes. Implications for finding sequence motifs in regulatory regions.
796 *Nucleic Acids Res* 29:2607–2618.
- 797 57. Delcher A, Bratke K, Powers E, Salzberg S. 2007. Identifying bacterial genes and
798 endosymbiont DNA with Glimmer. *Bioinformatics* 23:673–679.
- 799 58. Marchler-Bauer A, Lu S, Anderson JB, Chitsaz F, Derbyshire MK, DeWeese-Scott C,
800 Fong JH, Geer LY, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Jackson JD, Ke Z,
801 Lanczycki CJ, Lu F, Marchler GH, Mullokandov M, Omelchenko M V., Robertson CL,
802 Song JS, Thanki N, Yamashita RA, Zhang D, Zhang N, Zheng C, Bryant SH. 2011. CDD:
803 A Conserved Domain Database for the functional annotation of proteins. *Nucleic Acids*
804 *Res* 39:225–229.
- 805 59. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski
806 B, Ideker T. 2003. Cytoscape: A Software Environment for Integrated Models. *Genome*
807 *Res* 13:2498–2504.
- 808 60. Lederberg EM, Lederberg J. 1953. Genetic Studies of Lysogenicity in *Escherichia coli*.
809 *Genetics* 38:51–64.
- 810 61. Deatherage DE, Barrick JE. 2014. Identification of mutations in laboratory evolved
811 microbes from next-generation sequencing data using breseq. *Methods Mol Biol*
812 https://doi.org/10.1007/978-1-4939-0554-6_12.
- 813 62. Seemann T. 2015. snippy: fast bacterial variant calling from NGS reads. Retrieved from
814 <https://github.com/tseemann/snippy>.

- 815 63. Sambrook J, Russell DW. 2001. Molecular cloning: a laboratory manual 3rd edition.
816 Coldspring Harbour Laboratory Press, UK.
- 817 64. Silhavy TJ, Berman ML, Enquist LW. 1984. Experiments with gene fusions. Cold Spring
818 Harb Lab.
- 819 65. Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, Datsenko KA, Tomita M,
820 Wanner BL, Mori H. 2006. Construction of Escherichia coli K-12 in-frame, single-gene
821 knockout mutants: The Keio collection. Mol Syst Biol
822 <https://doi.org/10.1038/msb4100050>.
- 823 66. Datsenko KA, Wanner BL. 2000. One-step inactivation of chromosomal genes in
824 Escherichia coli K-12 using PCR products. Proc Natl Acad Sci U S A 97:6640–6645.
- 825 67. Hitchcock PJ. 1984. Analyses of gonococcal lipopolysaccharide in whole-cell lysates by
826 sodium dodecyl sulfate-polyacrylamide gel electrophoresis: Stable association of
827 lipopolysaccharide with the major outer membrane protein (protein I) of Neisseria
828 gonorrhoeae. Infect Immun 46:202–212.
- 829 68. Kulikov EE, Golomidova AK, Prokhorov NS, Ivanov PA, Letarov A V. 2019. High-
830 throughput LPS profiling as a tool for revealing of bacteriophage infection strategies. Sci
831 Rep 9.
- 832 69. Tsai C, Frasch C. 1982. Silver Stain for Detecting Lipopolysaccharides Polyacrylamide
833 Gels. Anal Biochem 19:115–119.
- 834 70. Davis MR, Goldberg JB. 2012. Purification and visualization of lipopolysaccharide from
835 gram-negative bacteria by hot aqueous-phenol extraction. J Vis Exp 1.

- 836 71. Wintersinger JA, Wasmuth JD. 2015. Kablammo: An interactive, web-based BLAST
837 results visualizer. *Bioinformatics* 31:1305–1306.
- 838 72. Guyer MS, Reed RR, Steitz JA, Low KB. 1981. Identification of a sex-factor-affinity site
839 in *E. coli* as gamma delta. *Cold Spring Harb Symp Quant Biol* 45:135–140.
- 840 73. Ju T, Shoblak Y, Gao Y, Yang K, Fouhse J, Finlay B, Wing So Y, Stothard P, Willing BP.
841 2017. Initial Gut Microbial Composition as a Key Factor Driving Host Response to
842 Antibiotic Treatment, as Exemplified by the Presence or Absence of Commensal
843 *Escherichia coli*. *Appl Environ Microbiol* 83.
- 844 74. Casadaban MJ. 1976. Transposition and fusion of the lac genes to selected promoters in
845 *Escherichia coli* using bacteriophage lambda and Mu. *J Mol Biol* 104:541–555.
- 846 75. Bachmann BJ. 1972. Pedigrees of some mutant strains of *Escherichia coli* K-12. *Bacteriol*
847 *Rev* 36:525–557.
- 848 76. Hull RA, Gill RE, Hsu P, Minshew BH, Falkow S. 1981. Construction and expression of
849 recombinant plasmids encoding type 1 or D-mannose-resistant pili from a urinary tract
850 infection *Escherichia coli* isolate. *Infect Immun* 33:933–938.
- 851 77. Iguchi A, Thomson NR, Ogura Y, Saunders D, Ooka T, Henderson IR, Harris D,
852 Asadulghani M, Kurokawa K, Dean P, Kenny B, Quail MA, Thurston S, Dougan G,
853 Hayashi T, Parkhill J, Frankel G. 2009. Complete genome sequence and comparative
854 genome analysis of enteropathogenic *Escherichia coli* O127:H6 strain E2348/69. *J*
855 *Bacteriol* 91:347–354.
- 856 78. Nissile A. 1918. Die antagonistische Behandlung chronischer Darmstörungen mit

857 Colibakterien. Med Klin 29–30.

858 79. Schauer DB, Falkow S. 1993. The *eae* gene of *Citrobacter freundii* biotype 4280 is
859 necessary for colonization in transmissible murine colonic hyperplasia. *Infect Immun*
860 61:4654–4661.

861

862 **Figure Legends**

863

864 **Figure 1. PHASTER analysis of MP1 and transmission electron microscopy image of Kapi1.**

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

871

872 **Figure 2. Genome of Kapi1.** (A) Layout of the Kapi1 genome, with ORFs color coded as follows:

873 lysis module in orange, structural/morphogenesis in blue, DNA replication/repair/regulators in
874 green, tRNAs in yellow, and hypothetical proteins and proteins of unknown function in grey.
875 Prophage_left and _right primers used to verify genome circularity are indicated in purple. (B)
876 Alignment of Kapi1 genome with top NCBI BLASTn hits. The genome of vB_EcoP_Kapi1
877 (accession MT813197) was searched against the NCBI Blast nucleotide database, and the top hits

878 were visualized with Kablammo (71); darker lines represent higher sequence homology. The upper
879 panel shows the alignment of Kapi1 with the top hit from NCBI database without any filters, while
880 the lower panel shows the top hit from NCBI database with the viral (taxid: 10239) filter applied.
881 (C) vContact2 (17) clustering of Kapi1-related viruses, visualized with CytoScape (59). The first-
882 neighbors network of Kapi1 was filtered to include only the top 50% strongest interactions with
883 Kapi1, and with edges that do not directly connect to Kapi1 were removed for clearer visualization.
884 An edge-weighted spring-embedded model was used so that the distance between viral nodes, and
885 the darkness and thickness of edges connecting the nodes corresponds to the similarity between
886 those viruses. Viral clusters are color-coded.

887

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

895

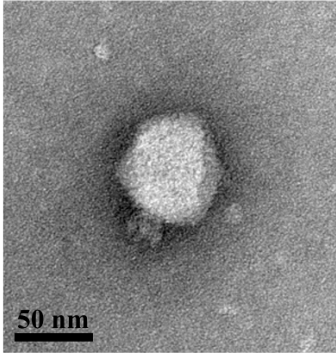
896 **Figure 4. The lifestyle of Kapi1.** (A) prophage liberation assay; cultures of MP13 (Kapi1 lysogen)
897 were grown in LB, 50% LB 50% simulated gastric fluid (SGF) (29), 50% LB 50% simulated
898 intestinal fluid (SIF) (29), or LB supplemented with 0.5 ng/uL mitomycin C for 24 hours. After 24
899 hours the number of cells were enumerated by spotting on LB plates, and the number of phages

900 were enumerated by spotting on soft agar overlays prepared with KP121. PFU/mL was divided by
901 CFU/mL to obtain PFU/CFU; the values represent the average of three biological replicates, and
902 error bars represent the standard deviation. A one-way ANOVA was performed with a Tukey's
903 post-hoc test; the number of phage released per cell in the LB, SGF, and SIF treatments varied
904 significantly from the mitomycin C treatment ($p < 0.0005$), but not from each other. (B) KP7 and
905 MP13 growth curve. KP7 and MP13 were grown overnight in biological triplicates, OD^{600} was
906 standardized to 1.0, then each strain was sub-cultured 1:100 into LB in 9 technical replicates in a
907 96-well plate. The plate was incubated in a plate-reader for 15.5 hours at 37 °C with OD^{600}
908 measurements taken every 30 minutes, preceded by a brief shaking interval. Values plotted are the
909 mean of 27 replicates, (excluding 2 outliers for KP7 which did not grow), with the standard
910 deviation shown as error bars. (C) Prophage stability assay; MP13 was grown in LB, 50% LB 50%
911 SGF (29), 50% LB 50% SIF (29), or LB supplemented with 0.5 ng/uL mitomycin C for 24 hours,
912 then sub-cultured into fresh media and grown for another 24 hours. After 48 hours total incubation,
913 cells were plated onto LB agar, then replica plated onto a second LB agar plate spread with KP7
914 (not lysogenic for Kapi1). The number of non-lysogenic colonies was divided by the total number
915 of colonies to obtain % loss of Kapi1; values represent the average of three biological replicates,
916 and error bars represent the standard deviation. A one-way ANOVA was performed on arcsin-
917 transformed values with a Tukey post-hoc test; the number of Kapi1 lysogens lost in the mitomycin
918 C treatment differed significantly from the LB treatment, and from the SIF treatment ($P < 0.01$).

919

920 **Figure 5. Kapi1 uses the O-antigen as a receptor and modifies its structure.** (A) LPS
921 extracted by proteinase K digestion was run on SDS-PAGE and silver stained; lane 1 – WT KP7,
922 lane 2 – MP13 Kap1 lysogen, lane 3 – KP61, lane 4 – KP62. (B) Schematic representing the

923 truncated *wzy* polymerase from KP61 and KP62 compared with the full-length WT protein. (C)
924 Kapi1 phage lysate was serially diluted and spotted onto soft-agar overlays with a variety of
925 mutants in the KP7 background. (D) Phenol-extracted LPS was run on SDS-PAGE and silver
926 stained; lane 1 – KP7, lane 2 – KP7 LPS treated with Kapi1 for 30 minutes, lane 3 – MP13.
927

A	Prophage Region	Most Common Phage	Accession	B
	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 500465-1	NC_049342	
	Incomplete_3	Bacillus phage G	NC_023719	

928

929 **Figure 1. PHASTER analysis of MP1 and transmission electron microscopy image of Kapi1.**

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

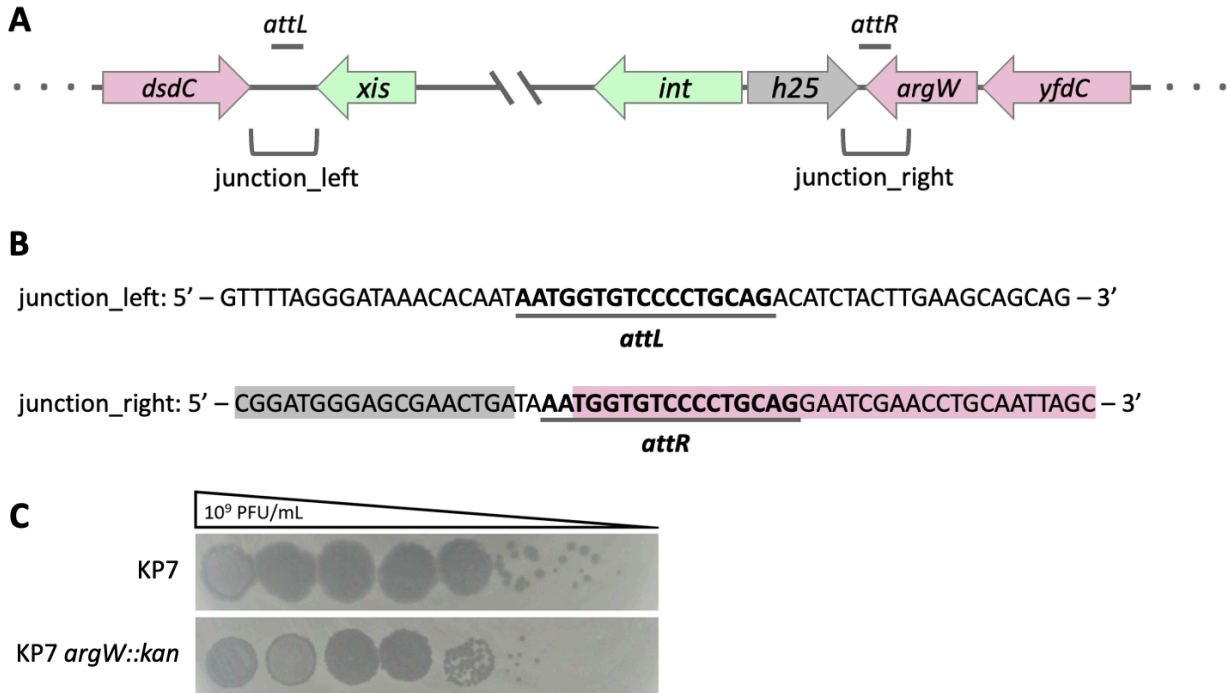
936

943 (accession MT813197) was searched against the NCBI Blast nucleotide database, and the top hits
944 were visualized with Kablammo (71); darker lines represent higher sequence homology. The upper
945 panel shows the alignment of Kapi1 with the top hit from NCBI database without any filters, while
946 the lower panel shows the top hit from NCBI database with the viral (taxid: 10239) filter applied.
947 (C) vContact2 (17) clustering of Kapi1-related viruses, visualized with CytoScape (59). The first-
948 neighbors network of Kapi1 was filtered to include only the top 50% strongest interactions with
949 Kapi1, and with edges that do not directly connect to Kapi1 were removed for clearer visualization.
950 An edge-weighted spring-embedded model was used so that the distance between viral nodes, and
951 the darkness and thickness of edges connecting the nodes corresponds to the similarity between
952 those viruses. Viral clusters are color-coded.

953

954

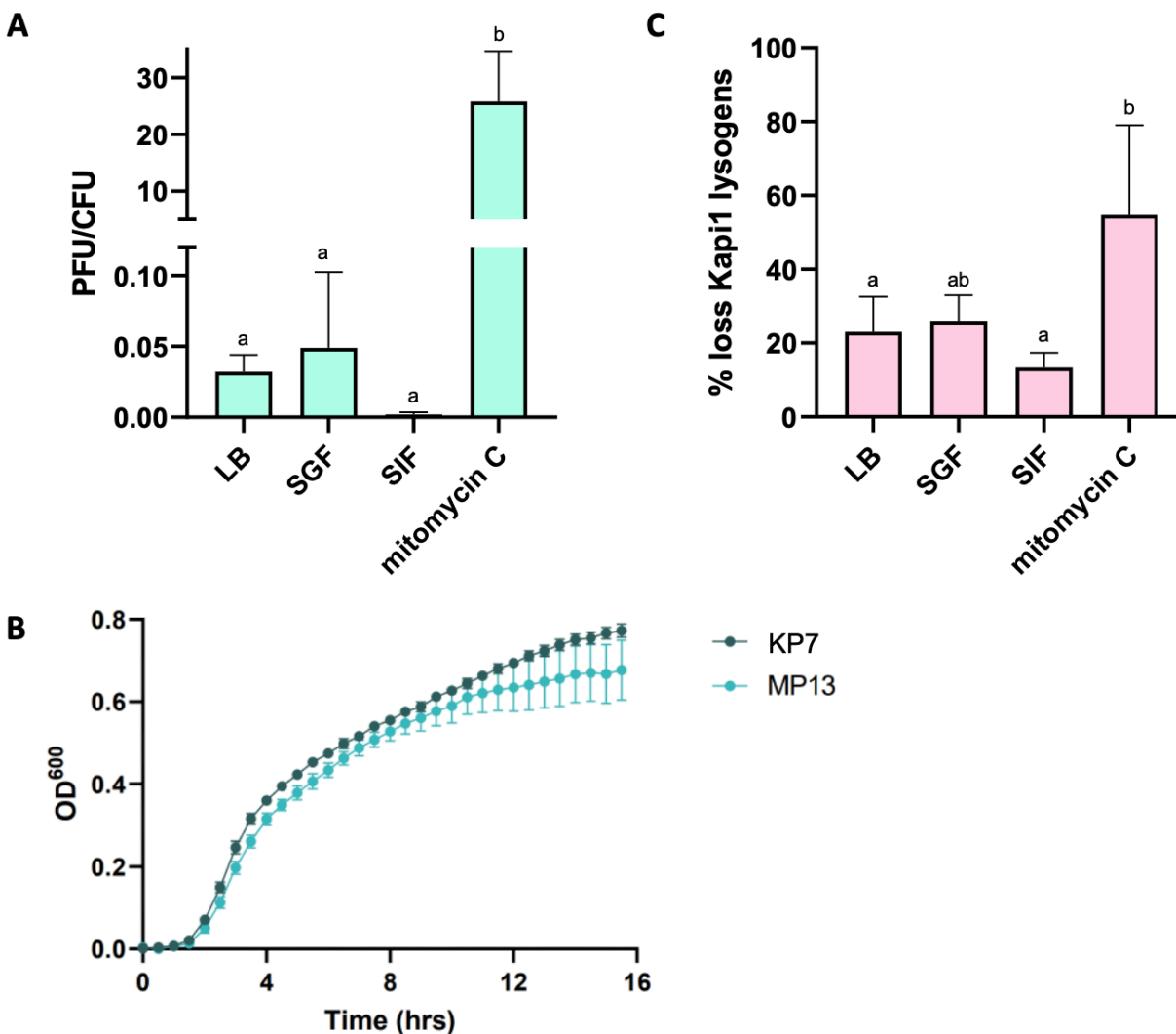
955



956

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

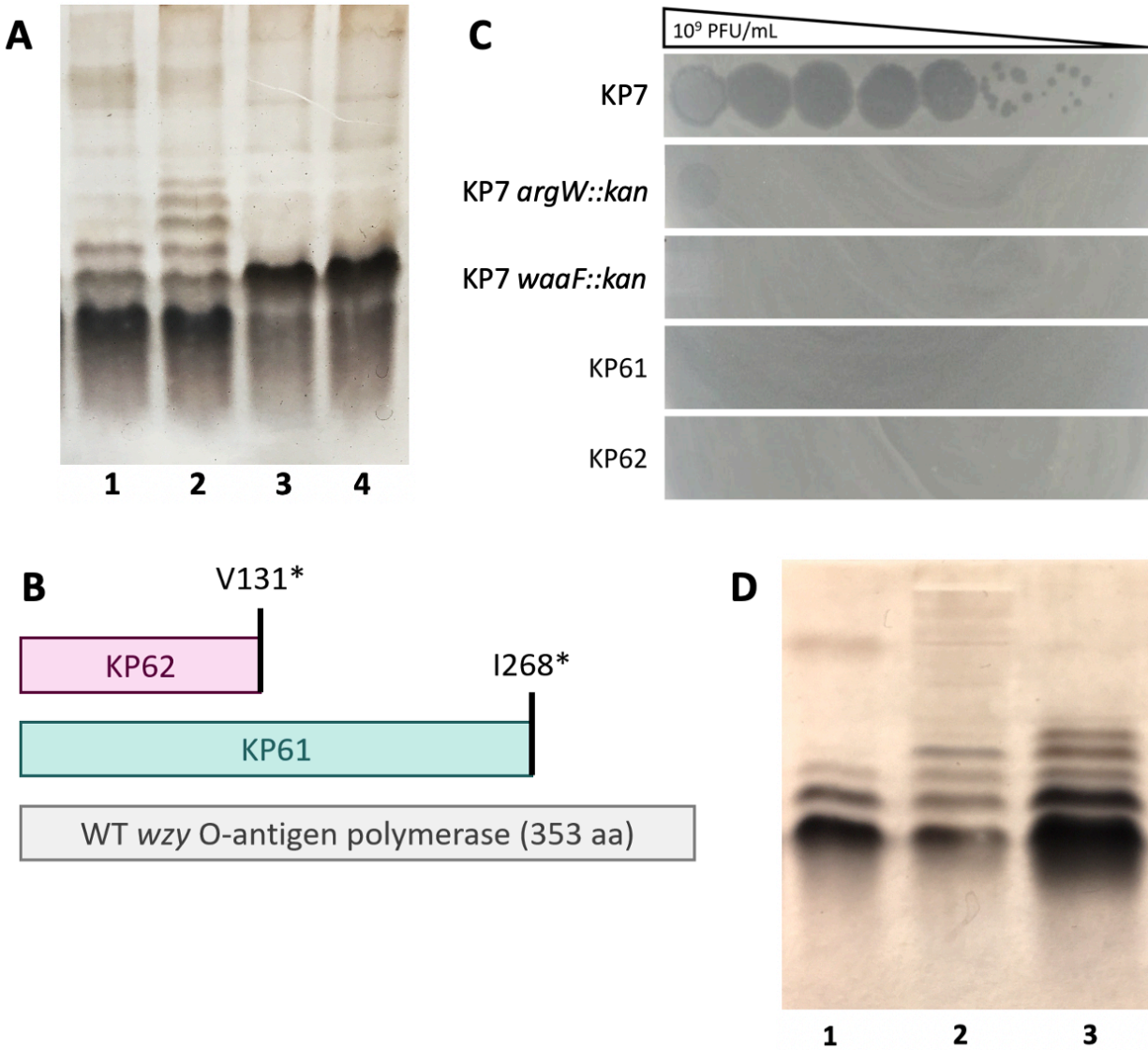
964



965

966 **Figure 4. The lifestyle of Kapi1.** (A) prophage liberation assay; cultures of MP13 (Kapi1 lysogen)
967 were grown in LB, 50% LB 50% simulated gastric fluid (SGF) (29), 50% LB 50% simulated
968 intestinal fluid (SIF) (29), or LB supplemented with 0.5 ng/uL mitomycin C for 24 hours. After 24
969 hours the number of cells were enumerated by spotting on LB plates, and the number of phages
970 were enumerated by spotting on soft agar overlays prepared with KP121. PFU/mL was divided by
971 CFU/mL to obtain PFU/CFU; the values represent the average of three biological replicates, and
972 error bars represent the standard deviation. A one-way ANOVA was performed with a Tukey's
973 post-hoc test; the number of phage released per cell in the LB, SGF, and SIF treatments varied

974 significantly from the mitomycin C treatment ($p < 0.0005$), but not from each other. (B) KP7 and
975 MP13 growth curve. KP7 and MP13 were grown overnight in biological triplicates, OD^{600} was
976 standardized to 1.0, then each strain was sub-cultured 1:100 into LB in 9 technical replicates in a
977 96-well plate. The plate was incubated in a plate-reader for 15.5 hours at 37 °C with OD^{600}
978 measurements taken every 30 minutes, preceded by a brief shaking interval. Values plotted are the
979 mean of 27 replicates, (excluding 2 outliers for KP7 which did not grow), with the standard
980 deviation shown as error bars. (C) Prophage stability assay; MP13 was grown in LB, 50% LB 50%
981 SGF (29), 50% LB 50% SIF (29), or LB supplemented with 0.5 ng/uL mitomycin C for 24 hours,
982 then sub-cultured into fresh media and grown for another 24 hours. After 48 hours total incubation,
983 cells were plated onto LB agar, then replica plated onto a second LB agar plate spread with KP7
984 (not lysogenic for Kapi1). The number of non-lysogenic colonies was divided by the total number
985 of colonies to obtain % loss of Kapi1; values represent the average of three biological replicates,
986 and error bars represent the standard deviation. A one-way ANOVA was performed on arcsin-
987 transformed values with a Tukey post-hoc test; the number of Kapi1 lysogens lost in the mitomycin
988 C treatment differed significantly from the LB treatment, and from the SIF treatment ($P < 0.01$).
989



990

991 **Figure 5. Kapi1 uses the O-antigen as a receptor and modifies its structure.** (A) LPS
992 extracted by proteinase K digestion was run on SDS-PAGE and silver stained; lane 1 – WT KP7,
993 lane 2 – MP13 Kap1 lysogen, lane 3 – KP61, lane 4 – KP62. (B) Schematic representing the
994 truncated *wzy* polymerase from KP61 and KP62 compared with the full-length WT protein. (C)
995 Kapi1 phage lysate was serially diluted and spotted onto soft-agar overlays with a variety of
996 mutants in the KP7 background. (D) Phenol-extracted LPS was run on SDS-PAGE and silver
997 stained; lane 1 – KP7, lane 2 – KP7 LPS treated with Kapi1 for 30 minutes, lane 3 – MP13.