Prophage elements function as reservoir for antibiotic resistance and virulence genes in nosocomial pathogens

Kohei Kondo[a,#], Mitsuoki Kawano[b], Motoyuki Sugai[a]

[a]Antimicrobial Resistance Research Center, National Institute of Infectious Diseases, Higashi Murayama, Tokyo, Japan

[b]Department of Human Nutrition, Faculty of Contemporary Life Science, Chugokugakuen University, Kita-ku, Okayama, Japan

Running head: AMR genes in prophages of ESKAPE pathogens and *E. coli*

#Correspondence to: kondo-k@niid.go.jp

Abstract

Prophages are often involved in host survival strategies and contribute toward increasing the genetic diversity of the host genome. Prophages also drive horizontal propagation of various genes as vehicles. However, there are few retrospective studies contributing to the propagation of antimicrobial resistance (AMR) and virulence factor (VF) genes by prophage. In this study, we extracted complete genome sequences of seven pathogens, including ESKAPE bacteria and *Escherichia coli* deposited in a public database, and examined the distribution of both AMR and VF genes in certain genomic regions of prophage, including prophage-like element. We found that the ratios of AMR and VF genes greatly varied among the seven species. More than 55% of *Enterobacter cloacae* strains had VF genes, but only 0.8% of *Klebsiella pneumoniae* strains had VF genes from prophages. The prophage types carrying AMR genes were detected in a broad range of hosts, whereas prophages containing VF genes were conserved in only one or two species, suggesting that distribution patterns of prophages were different between prophages encoding AMR or VF genes. We also found that the prophage containing class 1 integrase possessed a significantly higher number of AMR genes than prophages with no class 1 integrase. Moreover, AMR genes in the prophage were located near transposase and integrase. The results of this study reveal a comprehensive picture of AMR and VF genes present in prophage elements and provide new insights into the horizontal transfer of genes associated with antimicrobial resistance and pathogenicity.

Importance

Although we believe phages play an important role in horizontal gene transfer in exchanging genetic material, we do not know the distribution of the antimicrobial resistance and/or virulence genes in prophages. We collected different prophage elements from the complete genome sequence of seven species – *Enterococcus faecium*, *Staphylococcus aureus*, *Klebsiella pneumoniae*, *Acinetobacter baumannii, Pseudomonas aeruginosa*, and *Enterobacter cloacae*, as well as *Escherichia coli* –, and characterized the distribution of antimicrobial resistance and virulence genes encoded in the prophage region. While virulence genes in prophage were found to be species-specific, antimicrobial resistance genes in prophages were highly conserved in various species. Integron structure was detected within prophage regions in almost all of the genera. Maximum of 11 antimicrobial resistance genes were found in a single prophage region, suggesting that prophages act as a reservoir for antimicrobial resistance genes. Our results highlight new insights on prophages as horizontal gene carriers.

Introduction

Antimicrobial resistance (AMR) is a global public health issue. In recent years, ESKAPE pathogens (*Enterococcus faecium*, *Staphylococcus aureus*, *Klebsiella pneumoniae*, *Acinetobacter baumannii, Pseudomonas aeruginosa*, and *Enterobacter* spp.) have become a threat since they are the leading cause of nosocomial infection and easily escape from authentic chemotherapy due to their antimicrobial-resistant phenotype and many countries have faced difficulties in controlling these pathogens (1, 2). In Japan, *Escherichia coli* has replaced *Staphylococcus aureus* as the primary pathogen isolated from clinical samples in hospitals since 2018, and isolation of third-generation cephalosporin- or quinolone-resistant *E. coli* continues to increase in Japan (https://janis.mhlw.go.jp). Moreover, the extended-spectrum of ß-lactamase-producing *E. coli* is spreading worldwide (3). These ESKAPE and *E. coli* are major AMR pathogens in nosocomial settings.

A bacteriophage (phage) is a virus that infects bacteria. As soon as a phage adsorbs to the host's cell wall, the phage genome is injected into the host cell. Temperate phages follow one of the two life cycles afterwards; the lysogenic cycle or the lytic cycle. In the lysogenic cycle, the phage genome is integrated into the host chromosome and it is called a prophage. In the lytic cycle, the prophage is induced to produce progeny phages in response to chemical or physical stressors (4, 5). Lysogenic phages or prophages are known to drive horizontal gene transfer (HGT) through transduction, but they also play an important role in increasing the

64 genetic diversity of the host (6–10). Furthermore, defective prophages, also known as
65 prophage-like elements, are stable in the host genome despite deleting most of the phage
66 genes (6) and are known to increase host survival by conferring resistance against various
67 stresses (11–13).

68 Plasmid conjugation has been well established as the major means of HGT of (AMR) genes
69 (14), but recent studies have shed light on the role of phages in HGT of AMR genes since
70 they are often encoded within the genome of the phage or prophage (15, 16). Metagenomic
71 analysis revealed that AMR genes, such as ß-lactamases, are found in the phage genome (17–
72 19). A clinical study reported that phages harboring AMR genes were identified in samples
73 from patients with cystic fibrosis (20). Costa et al. reported that many prophage regions
74 within the *A. baumannii* genome possessed several AMR and virulence factor (VF) genes
75 using a bioinformatics approach (21). These studies have implied that phages and prophages
76 probably transduce AMR genes more frequently than expected. However, the relationship
77 between prophages and AMR genes has not been fully explored.

78 Pathogenic or VF genes in prophages have been mainly studied in *S. aureus* (22, 23), *E. coli*
79 (24, 25), *Salmonella enterica* (26), and *Vibrio* spp. (27, 28), and their pathogenicity is
80 associated with VFs encoded by the prophages. Other reports revealed that the expression of
81 the Shiga toxin in *E. coli* (29) and mitogen factor in *Streptococcus canis* (30) depends on
82 prophage induction. The studies mentioned above indicate that it is important to investigate

the connection between host pathogenicity and their prophage to reveal their pathogenesis. However, there is very little known on the relationship between virulence and prophages in the rapidly emerging multidrug-resistant bacteria, such as ESKAPE pathogens and *E. coli.* This study aims to understand the distribution of AMR and VF genes encoded in prophages, including the intact region, prophage-like elements, and satellite prophages (31) comprised in the bacterial genome, and discover their specific structural genomic features beyond the genera. We focus on seven clinically important AMR pathogens including ESKAPE pathogens and *E. coli* and analyzed their complete genomes deposited in a database to mine the prophage structure.

Results

Comparison of host genome size and the number of prophages

We investigated the correlation between the host genome size and the number of prophage elements present. The Pearson's correlation coefficients (R-values) of each species were 0.57 for *A. baumannii*, 0.51 for *E. cloacae*, 0.76 for *E. coli*, 0.67 for *E. faecium*, 0.51 for *K. pneumoniae*, 0.67 for *P. aeruginosa*, and 0.63 for *S. aureus*. The R-values for all species were greater than 0.5, indicating that the number of prophages positively correlated with the host genome size (Fig. 1A).

Next, we analyzed the number of prophages and prophage-like elements in each species. The number of prophages in *E. coli* was significantly higher compared to that in other species (Fig. 1B) ($P < 0.001$): *E. coli* possessed a maximum of 24 prophages (Accession numbers: CP027459 and CP024618) in which you can find prophages harboring Shiga toxin genes, *stx2A* and *stx2B*.

Screening for prophage elements in plasmids yielded a few cases, as depicted in Fig. S1. Comparison of the numbers of prophages in plasmids indicated that most *K. pneumoniae* plasmids harbored at least one or more prophages or prophage-like elements, and some *S. aureus* plasmids possessed one prophage or prophage-like element. However, *A. baumannii* possessed no plasmids harboring prophage or prophage-like elements. The number of prophages harbored in *K. pneumoniae* plasmids was significantly higher compared with that in *A. baumannii* and *S. aureus* plasmids ($p = 4.7 \times 10^{-5}$ and $p = 0.0011$, respectively) (Fig. S1). AMR genes from prophage elements were encoded in plasmids found in *E. faecium* (LC495616) and *K. pneumoniae* (AP018748, AP018752, AP018755 and AP018673), whereas no VF genes were detected on plasmid-harboring prophages in the accession numbers used in this experiment.

The proportion of the number of AMR or VF genes to that of prophages in the host genome

120 We next examined the proportion of prophage elements encoding either AMR or VF genes.
121 The proportions of prophages encoding VF genes were relatively high in *E. cloacae*, *E. coli*,
122 *E. faecium,* and *S. aureus* (74.1, 65.4, 46.5, and 88.8%, respectively) (Fig. 2). In contrast, no
123 prophages with VFs were detected in *A. baumannii*. Similarly, *K. pneumoniae* and *P.*
124 *aeruginosa* also showed a low proportion of VFs harbored in corresponding prophages (0.8
125 and 11.6%, respectively) (Fig. 2).
126 Prophage regions encoding AMR genes were detected in all species. The proportions were
127 21.2% for *A. baumannii*, 18.5% for *E. cloacae*, 13.3% for *E. coli*, 5.8% for *E. faecium*, 24.7%
128 for *K. pneumoniae*, 19.4% for *P. aeruginosa*, and 14% for *S. aureus*. The proportion of the
129 host genome that contained both AMR and VF genes in prophages was 14.8% for *E. cloacae*,
130 11.1% for *E. coli*, 2.3% for *E. faecium*, 0.3% for *K. pneumoniae*, 1.7% for *P. aeruginosa*, and
131 12.3% for *S. aureus* (Fig. 2). Overall, the differences in proportions for each species indicated
132 that the ratio between AMR and VF genes differed greatly depending on the species.
133
134 **Different prophage types encode either AMR or VF genes**
135 We investigated the phage types integrated into the genome in each species. The name of
136 each phage type was described using the most common phage from the PHASTER database.
137 Phage types carrying AMR gene(s) are listed in Fig. 3 and aligned according to the number of
138 phage types harboring AMR gene(s) (Fig. 3). The number of phage types harboring AMR

139 genes was 46 (Fig. 3). Notably, the major phage types carrying AMR gene(s), Escher_RCS47
140 (RCS47), Entero_phi80, and Salmon_SJ46, were present in multiple species beyond the
141 generic barrier. The Escher_RSC47 phage type was detected in all species except for *S.*
142 *aureus*, and the Staphy_SPbeta_like phage type was detected in all species except for *E.*
143 *cloacae* (Fig. 3).

144 In contrast, we found that VF genes were widely distributed in a large number of phage
145 types. There were 80 phage types, which were 1.7-fold higher than those harboring AMR
146 genes. Furthermore, the prophage types carrying VF genes were different from those carrying
147 AMR genes. Most of the phage types carrying VF genes, including ENT47670, phiFL1A,
148 and Diva, were not detected in the prophages carrying AMR genes (Fig. 3). In addition,
149 widely distributed VF-encoding prophages were not observed in various species. Overall,
150 these results indicated that AMR and VF genes rarely coexisted within the same prophage
151 and distribution pattern of prophage types containing AMR genes were different from that of
152 VF-encoding prophage types.

153

154 **Comparing complete genomes of prophages harboring AMR and VF genes**

155 We investigated and compared the completeness of prophage-encoding regions carrying
156 AMR and VF genes. The regions were classified using criteria from PHASTER as follows;
157 (1) intact, (2) questionable, or (3) incomplete. As shown in Fig. 4, percentages of prophages

158 with AMR genes were 30.4%, 43.7%, and 27.1% for incomplete, intact, and questionable
159 ones, respectively. Prophages carrying VF genes were mostly intact (70.3%), which indicated
160 that the prophage retained most of its region, whereas 19.4 and 10.7% were incomplete and
161 questionable phages, respectively (Fig. 4). In contrast to prophages carrying AMR genes,
162 those carrying VF genes often encoded proteins that were crucial for the structure of the
163 phage, such as the head, tail, and baseplate (Data Set S6 and S7). These phages were likely to
164 produce progeny phages infecting other hosts. All prophages encoding Shiga toxins in *E. coli*
165 were intact and contained phage structural proteins (Data Set S5). These data suggested that
166 VF genes in prophages, such as the Shiga toxin gene could be easily transduced by a
167 prophage.
168
169 **Characterization of VF genes in prophages**
170 To further investigate VF genes in prophage elements, VF genes for each accession number
171 were examined using the VFDB from ABRicate and the presence of VF genes in prophages
172 was visualized in the matrix (Fig. 5).
173 Prevalence of prophage-encoded VF genes was unique to each species (Fig. 5). For
174 example, we detected *gtrA* and *gtrB* in *E. cloacae* and *E. coli* strains and in just one *K.*
175 *pneumoniae* strain but not in other species (Fig. 5). Bacteriophage ENT47670 (accession
176 number: NC_019927) contained *gtrA* and *gtrB*. Interestingly, intact ENT47670 prophage,

177 which also contained structural phage proteins, was detected in the prophage region where
178 *gtrA* and *gtrB* were harbored (Data Set S5). This result suggested that *gtrA* and *gtrB* of the
179 prophage region in *E. cloacae* were acquired through the transduction of ENT47670 or
180 ENT47670-like phages (32) (Fig. 5; Data Set S5).
181 Most *E. faecium* had two specific genes in the prophage regions; *bsh* encoding a bile salt
182 hydrolase and *clpP* encoding an ATP-dependent protease, but other species did not (Fig. 5).
183 We mapped the prophage region harboring *bsh* and *clpP* in *E. faecium* using Easyfig (Fig.
184 S2). We found that these genes were generally located at or near the gaps in the genome, and
185 were scattered at various locations in the chromosome. These data suggested that coding
186 regions *bsh* and/or *clpP* in prophages or prophage-like elements were successfully acquired
187 via HGT. Classification of VF genes encoded by prophages revealed via VFDB keywords
188 revealed that each species possesses functionally unique VF genes (Fig. S3).
189
190 **Characterization of AMR genes encoded by or nearby prophage elements**
191 To investigate AMR gene distribution in the prophage elements of each strain in detail,
192 AMR genes in the prophage region for each accession number were extracted using the
193 ResFinder database from ABRicate and classified based on predicted substrates and mutation
194 locations (See material and methods). Predicted substrates were described in Data Set S2. All

195 species appeared to carry aminoglycoside modifying enzyme gene(s) and β-lactamase
196 encoding gene(s) except for *E. faecium* (Fig. 6).
197 In contrast, species-specific AMR genes were also detected. For instance, *A. baumannii*
198 harbored *msrE*, encoding the subunit for the ABC transporter and conferring resistance to
199 erythromycin and streptomycin. *vanHBX*, conferring resistance to vancomycin, was detected
200 in only *E. faecium*, and *vanHBX* was harbored in "intact" prophages with the genes encoding
201 some structural proteins (Data Set S4). This data suggested that a part of *van* family of genes
202 in vancomycin-resistant *E. faecium* were acquired via phage transduction. $bla_{\text{CTX-M}}$ (33) was
203 detected in the prophage region of several *K. pneumoniae* strains and only one *E. coli* strain
204 (Fig. 6).
205 AMR gene cassettes were widely detected in various species. For instance, *K. pneumoniae*
206 contained a cassette array of AMR genes including *sul1*, *aadA,* and *aacA* (Fig. 6). To detect
207 the combination or the gene cassette of AMR genes harbored by prophage elements, we
208 classified and clustered AMR gene cassettes for each prophage type (Fig. 7). Escher_RCS47,
209 Salmon_SJ46, Acineto_vB_AbaM_ME3, and Staphy_SPbeta_like harbored various AMR
210 genes (Fig. 7). In contrast, regions from prophages Entero_phi80, Entero_P1, and
211 Escher_HK639 harbored a set of AMR genes including *sul1_5*, *aac (3)_lb_1*, and *aadA2_1*
212 (Fig. 7). Another cassette array containing *ant(3'')-la_1*, *aph(3'')-la_7*, and *blaTEM-1D_1*
213 was detected in Escher_RCS47, Staphy_SPbeta_like, and Salmon_SJ46. Since these AMR

gene-combinations detected on prophage regions resemble the integron cassette array, we tried to identify the integrons in these prophage regions using the INTEGRALL database. We found that prophage regions in all species, except for *E. faecium* and *S. aureus,* possessed class 1 integrase. Therefore, we considered these characteristic regions containing AMR genes cassette array as an integron cassette array (Data Set S4).

Next, we wondered if the prophage region containing integrons had a higher number of AMR genes than regions without integrons (Fig. 8). We examined the number of AMR genes in phage regions carrying integrons (Int/P, red), in those without integrons (No-int, green), and in strains where integrons were present somewhere other than the prophage region (No-Int/P, blue).

Most of the phage regions (97%) carrying an integron possessed three or more AMR genes, whereas nearly 55% of phage regions without an integron possessed fewer AMR genes (Fig. 8A). These results indicated that the number of AMR genes were significantly higher in prophages with integrons than that in other groups (Fig. 8B).

Structural features of prophage elements containing AMR genes

We briefly overviewed structural features of the prophage region and its association with AMR gene(s) by showing representative prophage sequences (Fig. 9). The prophage regions within the attachment sites (*attL* and *attR*) are shown in Figure 9. AMR genes in intact

prophage regions existed either between or near integrase and/or transposase (Fig. 9A and B). Furthermore, AMR genes were located at the end of the prophage region, whereas the central position in the prophage region often encoded essential phage genes, especially structural proteins. Salmon_Fels_2, intact phage detected using PHASTER, lacked class 1 integron-integrase but phage-derived integrase was present just next to AMR gene(s) (Fig. 9A). We next visualized the prophage regions from Acineto_vB_AbaM_ME3, which were designated as incomplete or questionable phages using PHASTER and were termed as defective prophages (Fig. 9B). The prophage region from Acineto_vB_AbaM_ME3 had two different gene cassettes containing class 1 integrase and transposase, and both were located at the end of the region. The Acineto_vB_AbaM_ME3 type prophage region in strain CP020603 carried 11 AMR genes, which was the highest number per prophage-like elements in this study (Fig. 9B; Data Set S6).

In contrast to phage type described above, representative Escher_RCS47, which were classified as either incomplete or questionable, were devoid of most phage genes and their phage-derived transposases, and integrases were randomly arranged. Furthermore, AMR genes were located not only at the end but also at various positions including the center of the prophage sequence (Fig. 9C). The Escher_RCS47 prophage region was conserved in various species, and the characteristics of this prophage region might imply that these prophages had been integrated into the host genome for a long time. These results suggested that even

252 though many phage structural genes had been deleted in evolutionary processes, the
253 important genes for host survival, such as AMR genes, might have accumulated in such a
254 conservative prophage region.
255
256 **Discussion**
257 In this study, we collected the prophage elements from a total of 1,623 complete genomes of
258 nosocomial AMR pathogens deposited in a public database and characterized AMR and VF
259 genes harbored. Due to multi-drug resistance, ESKAPE bacteria have become a threat to
260 global health, especially in elder patients. HGT agents, such as plasmids, prophages, and
261 transposons often carry AMR and VF genes and move from bacteria to bacteria. Phage or
262 prophage-carrying AMR genes are reported to be minor HGT agents (34, 35). In recent years,
263 however, several studies have reported that AMR genes are present in the phage or prophage
264 region (36–39), implying that they are associated with the prophage region more than
265 expected. Nevertheless, there have been few reports comprehensively identifying and
266 analyzing AMR/VF genes encoded in the prophage region of nosocomial AMR pathogens.
267 A previous representative metagenomic study has reported that phage regions encode almost
268 no AMR genes (34). However, our study uncovered that the proportion of bacteria carrying
269 AMR genes on prophage-like elements reached 10–20% in each species (Fig. 2). Our results
270 suggested that the proportion of AMR genes encoded by prophages and defective prophages

271 was higher than that of those encoded in the genome of phage particles and that AMR genes
272 were preferentially accumulated in the prophage sequence and were stably inherited in the
273 host genome (6). To date, plasmids are the most well studied HGT agent. Nonetheless, herein
274 we did not compare the AMR gene richness between plasmids and prophage regions. Several
275 prophage-like elements encoding AMR genes were inserted, and shared AMR genes in a
276 plasmid; thus, we solely focused on the prevalence of AMR genes encoded by prophage
277 regions, to shed light on previously overlooked contribution of prophages in AMR gene
278 propagation.
279 We found that 57.5% of the prophages encoding AMR genes were defective prophages
280 (defined as incomplete or questionable using PHASTER) (Fig. 4), and they lacked phage
281 structural proteins (Fig. 4). Even if the prophage regions with AMR genes were determined
282 as “intact”, these prophages often carried no structural genes (Data Set S4). One may
283 speculate that these defective prophages cannot produce progeny phages due to the loss of
284 structural proteins. However, Hayashi et al. has reported that several defective prophages
285 encoding Shiga toxins in the genome of Enterohemorrhagic *E. coli* (EHEC) O157:H7 can
286 produce progeny phages and that progeny phages can infect to other hosts (40). They have
287 considered that prophages can complement each gene required for progeny phage
288 propagation by sharing and using genes present in several prophage regions. Based on

289 Hayashi's report, we speculated that several defective prophages carrying AMR genes can
290 complement the genes from each prophage region to synthesize progeny phages.
291 Prophage induction is triggered mainly by DNA damage via SOS response, and the host is
292 lysed by the phage lysin. We found that more than 57.5% of AMR gene-harboring prophages
293 were defective, whereas intact prophage was dominant in VF gene-harboring prophages
294 (70.4%) (Fig. 4). A previous study has shown that prophages confer antibiotic resistance to
295 the host and result in an increase of host viability (41). In other words, the host takes the
296 advantage of AMR gene products of prophages for survival. A recent study has reported that
297 the induction frequency of prophage carrying AMR genes decreases under the presence of
298 antibiotics (42). As a result, prophages tend to become defective more frequently and are
299 inherited in the host genome (43). In contrast, it has been reported that VF genes encoded in
300 prophage can contribute to increasing phage infectivity by increasing the burst size and the
301 latent period (44). In addition, prophages often harbor the genes involved in superinfection
302 exclusion, which is a phenomenon that phage or prophage prevents infection by other phages
303 (45, 46), and such genes are associated with host virulence (44). Therefore, we speculated
304 that VF genes encoded in prophages have presumably more benefits for the prophage rather
305 than bacteria, and the selective pressure of becoming defective is hardly caused in VF gene-
306 encoding prophage. As a result, prophages containing VF genes tend to remain "intact" with
307 phage structural genes.

Our results showed that most VF genes on prophage element were species-specific (Fig. 5), presumably because the mechanism of virulence to the host, such as the entry into the mammalian body (cell), and the toxicity differs depending on the species. Thus, each VF gene, which is responsible for virulence, could be detected in a species-specific manner. In contrast, modes of action of antimicrobials are common among bacterial species, and AMR genes are mobilized among bacterial genera through horizontal transfer; thus, similar AMR genes encoded in prophage elements are conserved in various bacterial species.

It has been known that genes for aminoglycoside modification enzyme and ß-lactamase are widely conserved in various species (47, 48). In this study, these resistance genes were detected in or nearby prophages and prophage-like elements of almost all species, in accordance with previous studied (47, 48) (Fig. 6), suggesting that phages were related to HGT of these highly conserved AMR genes. A previous study showed that different subclasses of aminoglycoside nucleotidyltransferase (ANT) were detected in Gram-positive and -negative bacteria (49). While Gram-positive bacteria often possess ANT (4', 6', 9'), Gram-negative bacteria often have ANT (2", 3"). In this study, ANT (6, 9) were detected on the prophage of *E. faecium* and *S. aureus*, but were rarely detected in Gram-negative bacteria (Fig. 6) in agreement with previous reports (49).

Duplication of the same AMR genes was detected in a single prophage region in particular, two or more aminoglycoside resistance genes with different mutations or subtypes were

detected in a single prophage (Data Set S4). Duplication of AMR genes in prophage can increase the expression level and resistance to antibiotics (50). Furthermore, gene duplication or the increase of copy number is suggested to result in tolerance to mutations (51), accelerating evolution and genetic diversity (52). Thus, prophage element region may be a factory of producing AMR genes, which may result in the host rapidly becoming resistant to novel antibiotics.

We found that many prophages harboring AMR genes possessed an integron structure, and the number of AMR genes in integron-harboring prophages were significantly higher than that in prophage regions without an integron (Fig. 8). Some prophages with an integron have a characteristic structure with two different integron cassettes and a maximum of 11 AMR genes integrated into one prophage (Fig. 9). We proposed that the prophage region had a role as a reservoir for AMR genes. Integron structure was not detected in *E. faecium* and *S. aureus* prophage regions, and prophages without an integron structure often possessed a prophage-derived transposase near AMR gene(s). Interestingly, prophages RCS47, SJ46, and SPbeta-like phages, which have been detected in various species, harbor AMR genes and HGT-related genes in the original phage genome (53). Probably, phages carrying HGT-related genes could be key players to promote the transfer and accumulation of AMR genes. Even if the prophage region loses its infectivity due to genome incompleteness, AMR genes can be transferred by transposase (and integrase) derived from the phage. The Mu-like phage is

346 present on a plasmid and the transposon Tn21 encoded in the phage region of the plasmid
347 shared AMR genes (54). Overall, transposase derived from prophage has the potential to
348 transfer AMR genes.

349 This study has some limitations. Some samples were collected from identical places, such as
350 the same hospital or area, and the prophage regions of the genome from the same location
351 were similar. Thus, further studies should focus on the relationships between the location and
352 features of prophages. The sample number is uneven between each species. For example, the
353 sample number of *E. cloacae* was 27, while that of *S. aureus* was 408. Since most of the
354 complete genomes of *E. coli* are the EHEC, Stx (classified as "Toxin" in Fig. S3) and type III
355 secretion systems were detected using the VFDB database, in agreement with previous
356 reports (55, 56). Therefore, the distribution of VF genes and the percentage of Toxin and type
357 III system harbored by prophages may have been influenced by the *E. coli* EHEC data. To
358 analyze detailed and precise prophage sequence structure, we utilized a highly accurate
359 prophage sequence using the National Center for Biotechnology Information (NCBI) RefSeq
360 (57). We believe that our finding was endowed using not miscellaneous sequences but
361 RefSeq, which curated highly accurate sequences.

362 Overall, we comprehensively detected AMR and VF genes in a wide range of strains, and
363 our results will shed light on the important roles of phages as reservoirs and factors that
364 transfer AMR/VF genes. Further research is needed to elucidate the amount of AMR and VF

365 genes that are transferred to other strains via transduction in a clinical and natural
366 environment.
367
368 **Materials and methods**
369 **Data collection and prophage region detection**
370 We compiled complete genomes and RefSeq data of seven bacterial species (169 sequences
371 of *A. baumannii*, 27 sequences of *E. cloacae*, 324 sequences of *E. coli*, 88 sequences of *E.*
372 *faecium*, 408 sequences of *K. pneumoniae*, 183 sequences of *P. aeruginosa*, and 424
373 sequences of *S. aureus*) from GenBank using Biopython version 1.76 (58). The genome size
374 for each strain was referred to as the value described in GenBank (Data Set S1). Since *E. coli*
375 had a large number of registrations on the NCBI, GenBank accession numbers were
376 randomly selected so that the sample size would not increase. To detect prophages and
377 prophage-like elements (28) for each strain, we used a custom application programming
378 interface (API) from PHASTER (59). Prophage names were identified using the most
379 common phage species described in PHASTER. Additionally, PHASTER was also used to
380 classify prophages as either 1) “intact”, 2) “questionable”, or 3) “incomplete” based on the
381 length of the phage region and the number of phage-derived genes (Data Set S4 and S5).
382
383 **Detection of AMR and VF genes**

AMR and VF genes encoded by the prophage sequences were extracted using ABRicate version 1.0.1 (https://github.com/tseemann/abricate) under default settings. The ResFinder database (60) was used to detect AMR genes, and the Virulence Factors Database (VFDB) (61) was used to detect VF genes. VFs were classified based on VFDB keywords, and redundant (similar) keywords were summarized into short words (e.g., “adhesion,” “Apoptosis and Adherence,” and “Invasive” keywords were integrated into “Adherence,” and toxin-based genes were integrated into “toxin”). The same gene with a few, different mutations were distinguished as an accession number were assigned to each gene containing the respective mutation. Each of the genes was also assigned a "_ X" suffix.

Integron analysis

Integrons and antibiotic cassette gene arrays in prophages were detected using INTEGRALL (62). Ambiguous integrases encoded in prophage region were examined using Blastp (protein-protein BLAST). If the amino acid identity was more than 80%, it was considered as a class 1 integrase. Numbers were assigned based on the order of the gene cassette and we referred to the number described in INTEGRALL. We classified the results into three groups, depending on the type of integron arrangement, as follows: 1) Prophage region with a class 1 integrase (Int/P); 2) prophage region without a class 1 integrase (No-Int); 3) integrase present, but did not exist in the prophage region (No-Int/P).

403 **Prophage elements and other data visualization**

404 Prophage elements encoding *bsh* and *clpP*, were visualized using Blast Ring Image
405 Generator (BRIG) version 0.95 (63). All BRIG parameters were set to default. Other
406 prophage-related regions containing AMR genes were visualized using Easyfig version 2.2.2
407 (64). The thresholds of BLAST hits in Easyfig were analyzed using the default value. To
408 visualize, we selected Escher_HK639, Entero_186, and Salmon_Fels_2, which were regarded
409 as "intact" phages, encoding a phage structure protein. Alternatively, we selected
410 Acineto_vB_AbaM_ME3, Entero_phi80, and Escher_RCS47, which were considered
411 "incomplete" or "questionable" in PHASTER. The host accession number used for the
412 analysis had three to five strains selected at random for each prophage. Other data were
413 analyzed and visualized using Python 3.7.6 (Python Software Foundation,
414 https://www.python.org), Matplotlib version 3.1.3, and Seaborn version 0.1.0.

415

416 **Statistical analysis**

417 Pearson R correlation was calculated using the default jointplot function in seaborn. All
418 statistical analyses were conducted using a two-sided Welch's t-test using Python version
419 3.7.6 and SciPy Module version 1.4.1, and $P < 0.05$ was considered as a significant
420 difference.

421

Data availability

This study was performed using complete genomes registered on the NCBI. All accession numbers are listed in Data Set S1. All information on AMR and VF genes detected in this study are described in Data Set S2, S3. The information on prophages and AMR/VF gens including integron are summarized in Data Set S4 and S5. The presence of attachment sites (*attL*, *attR*) in each prophage region is described in Data Set S6, S7.

Acknowledgments

We are grateful to Professor Yoshitoshi Ogura for his valuable suggestions on the manuscript.

Reference

1. Rice LB. 2008. Federal Funding for the Study of Antimicrobial Resistance in Nosocomial Pathogens: No ESKAPE. J Infect Dis 197:1079–1081. https://doi.org/10.1086/533452.
2. Mulani MS, Kamble EE, Kumkar SN, Tawre MS, Pardesi KR. 2019. Emerging strategies to combat ESKAPE pathogens in the era of antimicrobial resistance: A review. Front Microbiol 10:539. https://doi.org/10.3389/fmicb.2019.00539.
3. Chong Y, Shimoda S, Shimono N. 2018. Current epidemiology, genetic evolution and clinical impact of extended-spectrum β-lactamase-producing *Escherichia col*i and

Klebsiella pneumoniae. Infect Genet Evol 61:185–188. https://doi.org/10.1016/j.meegid.2018.04.005.

4. Choi J, Kotay SM, Goel R. 2010. Various physico-chemical stress factors cause prophage induction in Nitrosospira multiformis 25196- an ammonia oxidizing bacteria. Water Res 44:4550–4558. https://doi.org/10.1016/j.watres.2010.04.040.

5. Boling L, Cuevas DA, Grasis JA, Kang HS, Knowles B, Levi K, Maughan H, McNair K, Rojas MI, Sanchez SE, Smurthwaite C, Rohwer F. 2020. Dietary prophage inducers and antimicrobials: toward landscaping the human gut microbiome. Gut Microbes 4:721–734. https://doi.org/10.1080/19490976.2019.1701353.

6. Bobay LM, Rocha EPC, Touchon M. 2013. The adaptation of temperate bacteriophages to their host genomes. Mol Biol Evol 30:737–751. https://doi.org/10.1093/molbev/mss279.

7. Croucher NJ, Coupland PG, Stevenson AE, Callendrello A, Bentley SD, Hanage WP. 2014. Diversification of bacterial genome content through distinct mechanisms over different timescales. Nat Commun 5:5471. https://doi.org/10.1038/ncomms6471.

8. Wyres KL, Wick RR, Judd LM, Froumine R, Tokolyi A, Gorrie CL, Lam MMC, Duchêne S, Jenney A, Holt KE. 2019. Distinct evolutionary dynamics of horizontal gene transfer in drug resistant and virulent clones of *Klebsiella pneumoniae*. PLoS Genet 15:e1008114. https://doi.org/10.1371/journal.pgen.1008114.

460 9. Shen J, Zhou J, Xu Y, Xiu Z. 2020. Prophages contribute to genome plasticity of
461 *Klebsiella pneumoniae* and may involve the chromosomal integration of ARGs in
462 CG258. Genomics 112:998–1010. https://doi.org/10.1016/j.ygeno.2019.06.016.
463 10. de Sousa JAM, Buffet A, Haudiquet M, Rocha EPC, Rendueles O. 2020. Modular
464 prophage interactions driven by capsule serotype select for capsule loss under phage
465 predation. ISME J 30:1–7. https://doi.org/10.1038/s41396-020-0726-z.
466 11. Silver-Mysliwiec TH, Bramucci MG. 1990. Bacteriophage-enhanced sporulation:
467 Comparison of spore-converting bacteriophages PMB12 and SP10. J Bacteriol
468 172:1948–1953. https://doi.org/10.1128/jb.172.4.1948-1953.1990.
469 12. McGrath S, Fitzgerald GF, Van Sinderen D. 2002. Identification and characterization
470 of phage-resistance genes in temperate lactococcal bacteriophages. Mol Microbiol
471 43:509–520. https://doi.org/10.1046/j.1365-2958.2002.02763.x.
472 13. Aucouturier A, Chain F, Langella P, Bidnenko E. 2018. Characterization of a
473 prophage-free derivative strain of *Lactococcus lactis* ssp. lactis IL1403 reveals the
474 importance of prophages for phenotypic plasticity of the host. Front Microbiol 9:2032.
475 https://doi.org/10.3389/fmicb.2018.02032.
476 14. Brown-Jaque M, Calero-Cáceres W, Muniesa M. 2015. Transfer of antibiotic-
477 resistance genes via phage-related mobile elements. Plasmid 79:1–7.
478 https://doi.org/10.1016/j.plasmid.2015.01.001.

15. Gómez-Gómez C, Blanco-Picazo P, Brown-Jaque M, Quirós P, Rodríguez-Rubio L, Cerdà-Cuellar M, Muniesa M. 2019. Infectious phage particles packaging antibiotic resistance genes found in meat products and chicken feces. Sci Rep 9:13281. https://doi.org/10.1038/s41598-019-49898-0.

16. López-Leal G, Santamaria RI, Cevallos MÁ, Gonzalez V, Castillo-Ramírez S. 2020. Prophages Encode Antibiotic Resistance Genes in *Acinetobacter baumannii*. Microb Drug Resist 00:10–12. https://doi.org/10.1089/mdr.2019.0362.

17. Anand T, Bera BC, Vaid RK, Barua S, Riyesh T, Virmani N, Hussain M, Singh RK, Tripathi BN. 2016. Abundance of antibiotic resistance genes in environmental bacteriophages. J Gen Virol 97:3458–3466. https://doi.org/10.1099/jgv.0.000639.

18. Rands CM, Starikova E V., Brüssow H, Kriventseva E V., Govorun VM, Zdobnov EM. 2018. ACI-1 beta-lactamase is widespread across human gut microbiomes due to transposons harboured by tailed prophages. Environ Microbiol 20:2288–2300. https://doi.org/10.1111/1462-2920.14276.

19. Moon K, Jeon JH, Kang I, Park KS, Lee K, Cha CJ, Lee SH, Cho JC. 2020. Freshwater viral metagenome reveals novel and functional phage-borne antibiotic resistance genes. Microbiome 8:75. https://doi.org/10.1186/s40168-020-00863-4.

20. Brown-Jaque M, Oyarzun LR, Cornejo-Sánchez T, Martín-Gómez MT, Gartner S, de Gracia J, Rovira S, Alvarez A, Jofre J, González-López JJ, Muniesa M. 2018.

498 Detection of bacteriophage particles containing antibiotic resistance genes in the
499 sputum of cystic fibrosis patients. Front Microbiol 9:856.
500 https://doi.org/10.3389/fmicb.2018.00856.
501 21. Costa AR, Monteiro R, Azeredo J. 2018. Genomic analysis of *Acinetobacter*
502 *baumannii* prophages reveals remarkable diversity and suggests profound impact on
503 bacterial virulence and fitness. Sci Rep 8:15346. https://doi.org/10.1038/s41598-018-
504 33800-5.
505 22. Bae T, Baba T, Hiramatsu K, Schneewind O. 2006. Prophages of Staphylococcus
506 aureus Newman and their contribution to virulence. Mol Microbiol 62:1035–1047.
507 https://doi.org/10.1111/j.1365-2958.2006.05441.x.
508 23. Dini M, Shokoohizadeh L, Jalilian FA, Moradi A, Arabestani MR. 2019. Genotyping
509 and characterization of prophage patterns in clinical isolates of *Staphylococcus aureus*.
510 BMC Res Notes 12:669. https://doi.org/10.1186/s13104-019-4711-4.
511 24. Hayashi T, Makino K, Ohnishi M, Kurokawa K, Ishii K, Yokoyama K, Han CG,
512 Ohtsubo E, Nakayama K, Murata T, Tanaka M, Tobe T, Iida T, Takami H, Honda T,
513 Sasakawa C, Ogasawara N, Yasunaga T, Kuhara S, Shiba T, Hattori M, Shinagawa H.
514 2001. Complete genome sequence of enterohemorrhagic *Eschelichia coli* O157:H7 and
515 genomic comparison with a laboratory strain K-12. DNA Res 8:11–22.
516 https://doi.org/10.1093/dnares/8.1.11.

517 25. Zhang Y, Liao Y Te, Salvador A, Sun X, Wu VCH. 2020. Prediction, Diversity, and
518 Genomic Analysis of Temperate Phages Induced From Shiga Toxin-Producing
519 *Escherichia coli* Strains. Front Microbiol 10:3093.
520 https://doi.org/10.3389/fmicb.2019.03093.
521 26. Cooke FJ, Wain J, Fookes M, Ivens A, Thomson N, Brown DJ, Threlfall EJ, Gunn G,
522 Foster G, Dougan G. 2007. Prophage sequences defining hot spots of genome variation
523 in *Salmonella enterica* serovar Typhimurium can be used to discriminate between field
524 isolates. J Clin Microbiol 45:2590–2598. https://doi.org/10.1128/JCM.00729-07.
525 27. Waldor MK, Mekalanos JJ. 1996. Lysogenic conversion by a filamentous phage
526 encoding cholera toxin. Science 272:1910–1913.
527 https://doi.org/10.1126/science.272.5270.1910.
528 28. Castillo D, Kauffman K, Hussain F, Kalatzis P, Rørbo N, Polz MF, Middelboe M.
529 2018. Widespread distribution of prophage-encoded virulence factors in marine Vibrio
530 communities. Sci Rep 8:9973. https://doi.org/10.1038/s41598-018-28326-9.
531 29. Kimmitt PT, Harwood CR, Barer MR. 1999. Induction of type 2 Shiga toxin synthesis
532 in *Escherichia coli* 0157 by 4-quinolones. Lancet 353:1588–1589.
533 https://doi.org/10.1016/S0140-6736(99)00621-2.

534 30. Ingrey KT, Ren J, Prescott JF. 2003. A fluoroquinolone induces a novel mitogen-
535 encoding bacteriophage in Streptococcus canis. Infect Immun 71:3028–3033.
536 https://doi.org/10.1128/IAI.71.6.3028-3033.2003.
537 31. Rezaei Javan R, Ramos-Sevillano E, Akter A, Brown J, Brueggemann AB. 2019.
538 Prophages and satellite prophages are widespread in Streptococcus and may play a role
539 in pneumococcal pathogenesis. Nat Commun 10:4852. https://doi.org/10.1038/s41467-
540 019-12825-y.
541 32. Li Y, Huang J, Wang X, Xu C, Han T, Guo X. 2020. Genetic Characterization of the
542 O-Antigen and Development of a Molecular Serotyping Scheme for Enterobacter
543 cloacae. Front Microbiol 11:727. https://doi.org/10.3389/fmicb.2020.00727.
544 33. Wyres KL, Holt KE. 2018. *Klebsiella pneumoniae* as a key trafficker of drug
545 resistance genes from environmental to clinically important bacteria. Curr Opin
546 Microbiol 45:131–139. https://doi.org/10.1016/j.mib.2018.04.004.
547 34. Enault F, Briet A, Bouteille L, Roux S, Sullivan MB, Petit MA. 2017. Phages rarely
548 encode antibiotic resistance genes: A cautionary tale for virome analyses. ISME J
549 11:237–247. https://doi.org/10.1038/ismej.2016.90.
550 35. Volkova V V., Lu Z, Besser T, Gröhn YT. 2014. Modeling the infection dynamics of
551 bacteriophages in enteric *Escherichia coli*: Estimating the contribution of transduction

552 to antimicrobial gene spread. Appl Environ Microbiol 80:4350–4362.
553 https://doi.org/10.1128/AEM.00446-14.
554 36. Modi SR, Lee HH, Spina CS, Collins JJ. 2013. Antibiotic treatment expands the
555 resistance reservoir and ecological network of the phage metagenome. Nature
556 499:219–222. https://doi.org/10.1038/nature12212.
557 37. Colomer-Lluch M, Calero-Cáceres W, Jebri S, Hmaied F, Muniesa M, Jofre J. 2014.
558 Antibiotic resistance genes in bacterial and bacteriophage fractions of Tunisian and
559 Spanish wastewaters as markers to compare the antibiotic resistance patterns in each
560 population. Environ Int 73:167–175. https://doi.org/10.1016/j.envint.2014.07.003.
561 38. Venturini C, Zingali T, Wyrsch ER, Bowring B, Iredell J, Partridge SR, Djordjevic SP.
562 2019. Diversity of P1 phage-like elements in multidrug resistant *Escherichia coli*. Sci
563 Rep 9:18861. https://doi.org/10.1038/s41598-019-54895-4.
564 39. Wang M, Zeng Z, Jiang F, Zheng Y, Shen H, Macedo N, Sun Y, Sahin O, Li G. 2020.
565 Role of enterotoxigenic *Escherichia coli* prophage in spreading antibiotic resistance in
566 a porcine-derived environment. Environ Microbiol https://doi.org/10.1111/1462-
567 2920.15084https://doi.org/10.1111/1462-2920.15084.
568 40. Asadulghani M, Ogura Y, Ooka T, Itoh T, Sawaguchi A, Iguchi A, Nakayama K,
569 Hayashi T. 2009. The defective prophage pool of *Escherichia coli* O157: Prophage-

570 prophage interactions potentiate horizontal transfer of virulence determinants. PLoS
571 Pathog 5:e1000408. https://doi.org/10.1371/journal.ppat.1000408.
572 41. Wang X, Kim Y, Ma Q, Hong SH, Pokusaeva K, Sturino JM, Wood TK. 2010. Cryptic
573 prophages help bacteria cope with adverse environments. Nat Commun 1:147.
574 https://doi.org/10.1038/ncomms1146.
575 42. Wendling C, Refardt D, Hall A. 2020. Fitness Benefits To Bacteria of Carrying
576 Prophages and Prophage-Encoded Antibiotic-Resistance Genes Peak in Different
577 Environments. bioRxiv
578 https://doi.org/10.1101/2020.03.13.990044https://doi.org/10.1101/2020.03.13.990044.
579 43. Bobay LM, Touchon M, Rocha EPC. 2014. Pervasive domestication of defective
580 prophages by bacteria. Proc Natl Acad Sci U S A 111:12127–12132.
581 https://doi.org/10.1073/pnas.1405336111.
582 44. Abedon ST, LeJeune JT. 2005. Why Bacteriophage Encode Exotoxins and other
583 Virulence Factors. Evol Bioinforma 1:97–110.
584 https://doi.org/10.1177/117693430500100001.
585 45. Bondy-Denomy J, Qian J, Westra ER, Buckling A, Guttman DS, Davidson AR,
586 Maxwell KL. 2016. Prophages mediate defense against phage infection through
587 diverse mechanisms. ISME J 10:2854–2866. https://doi.org/10.1038/ismej.2016.79.

588 46. Dedrick RM, Jacobs-Sera D, Guerrero Bustamante CA, Garlena RA, Mavrich TN,
589 Pope WH, Cervantes Reyes JC, Russell DA, Adair T, Alvey R, Bonilla JA, Bricker JS,
590 Brown BR, Byrnes D, Cresawn SG, Davis WB, Dickson LA, Edgington NP, Findley
591 AM, Golebiewska U, Grose JH, Hayes CF, Hughes LE, Hutchison KW, Isern S,
592 Johnson AA, Kenna MA, Klyczek KK, Mageeney CM, Michael SF, Molloy SD,
593 Montgomery MT, Neitzel J, Page ST, Pizzorno MC, Poxleitner MK, Rinehart CA,
594 Robinson CJ, Rubin MR, Teyim JN, Vazquez E, Ware VC, Washington J, Hatfull GF.
595 2017. Prophage-mediated defence against viral attack and viral counter-defence. Nat
596 Microbiol 2:16251. https://doi.org/10.1038/nmicrobiol.2016.251.
597 47. Fouhy F, Ogilvie LA, Jones B V., Ross RP, Ryan AC, Dempsey EM, Fitzgerald GF,
598 Stanton C, Cotter PD. 2014. Identification of aminoglycoside and β-lactam resistance
599 genes from within an infant gut functional metagenomic library. PLoS One 9:e108016.
600 https://doi.org/10.1371/journal.pone.0108016.
601 48. Ullmann IF, Tunsjø HS, Andreassen M, Nielsen KM, Lund V, Charnock C. 2019.
602 Detection of aminoglycoside resistant bacteria in sludge samples from Norwegian
603 drinking water treatment plants. Front Microbiol 10:487.
604 https://doi.org/10.3389/fmicb.2019.00487.
605 49. Ramirez MS, Tolmasky ME. 2010. Aminoglycoside modifying enzymes. Drug Resist
606 Updat 13:151–171. https://doi.org/10.1016/j.drup.2010.08.003.

607 50. Lee KI, Kusumoto M, Sekizuka T, Kuroda M, Uchida I, Iwata T, Okamoto S, Yabe K,
608 Inaoka T, Akiba M. 2015. Extensive amplification of GI-VII-6, a multidrug resistance
609 genomic island of *Salmonella enterica* serovar Typhimurium, increases resistance to
610 extended-spectrum cephalosporins. Front Microbiol 6:78.
611 https://doi.org/10.3389/fmicb.2015.00078.
612 51. Steenwyk JL, Rokas A. 2018. Copy number variation in fungi and its implications for
613 wine yeast genetic diversity and adaptation. Front Microbiol 9:288.
614 https://doi.org/10.3389/fmicb.2018.00288.
615 52. Force A, Lynch M, Pickett FB, Amores A, Yan YL, Postlethwait J. 1999. Preservation
616 of duplicate genes by complementary, degenerative mutations. Genetics 151:1531–
617 1545.
618 53. Billard-Pomares T, Fouteau S, Jacquet ME, Roche D, Barbe V, Castellanos M, Bouet
619 JY, Cruveiller S, Médigue C, Blanco J, Clermont O, Denamur E, Branger C. 2014.
620 Characterization of a P1-like bacteriophage carrying an SHV-2 extended-spectrum β-
621 lactamase from an *Escherichia coli* strain. Antimicrob Agents Chemother 58:6550–
622 6557. https://doi.org/10.1128/AAC.03183-14.
623 54. Radstrom P, Skold O, Swedberg G, Flensburg J, Roy PH, Sundstrom L. 1994.
624 Transposon Tn5090 of plasmid R751, which carries an integron, is related to Tn7, Mu,

and the retroelements. J Bacteriol 176:3257–3268. https://doi.org/10.1128/jb.176.11.3257-3268.1994.

55. Melton-celsa AR. 2014. Shiga Toxin (Stx) Classification, Structure, and Function. Microbiol Spectr 2:EHEC-0024–0013. https://doi.org/10.1128/microbiolspec.EHEC-0024-2013.Correspondence.

56. Jarvis KG, Girón JA, Jerse AE, Mcdaniel TK, Donnenberg MS, Kaper JB. 1995. Enteropathogenic *Escherichia coli* contains a putative type III secretion system necessary for the export of proteins involved in attaching and effacing lesion formation. Proc Natl Acad Sci U S A 92:7996–8000. https://doi.org/10.1073/pnas.92.17.7996.

57. d'Humières C, Touchon M, Dion S, Cury J, Ghozlane A, Garcia-Garcera M, Bouchier C, Ma L, Denamur E, P.C.Rocha E. 2019. A simple, reproducible and cost-effective procedure to analyse gut phageome: from phage isolation to bioinformatic approach. Sci Rep 9:11331. https://doi.org/10.1038/s41598-019-47656-w.

58. Cock PJA, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, Friedberg I, Hamelryck T, Kauff F, Wilczynski B, De Hoon MJL. 2009. Biopython: Freely available Python tools for computational molecular biology and bioinformatics. Bioinformatics 25:1422–1423. https://doi.org/10.1093/bioinformatics/btp163.

643 59. Arndt D, Grant JR, Marcu A, Sajed T, Pon A, Liang Y, Wishart DS. 2016. PHASTER:
644 a better, faster version of the PHAST phage search tool. Nucleic Acids Res 44:W16–
645 W21. https://doi.org/10.1093/nar/gkw387.
646 60. Zankari E, Hasman H, Cosentino S, Vestergaard M, Rasmussen S, Lund O, Aarestrup
647 FM, Larsen MV. 2012. Identification of acquired antimicrobial resistance genes. J
648 Antimicrob Chemother 67:2640–2644. https://doi.org/10.1093/jac/dks261.
649 61. Chen L, Zheng D, Liu B, Yang J, Jin Q. 2016. VFDB 2016: Hierarchical and refined
650 dataset for big data analysis - 10 years on. Nucleic Acids Res 44:D694–D697.
651 https://doi.org/10.1093/nar/gkv1239.
652 62. Moura A, Soares M, Pereira C, Leitão N, Henriques I, Correia A. 2009. INTEGRALL:
653 A database and search engine for integrons, integrases and gene cassettes.
654 Bioinformatics 25:1096–1098. https://doi.org/10.1093/bioinformatics/btp105.
655 63. Alikhan NF, Petty NK, Ben Zakour NL, Beatson SA. 2011. BLAST Ring Image
656 Generator (BRIG): Simple prokaryote genome comparisons. BMC Genomics 12:402.
657 https://doi.org/10.1186/1471-2164-12-402.
658 64. Sullivan MJ, Petty NK, Beatson SA. 2011. Easyfig: A genome comparison visualizer.
659 Bioinformatics 27:1009–1010. https://doi.org/10.1093/bioinformatics/btr039.
660

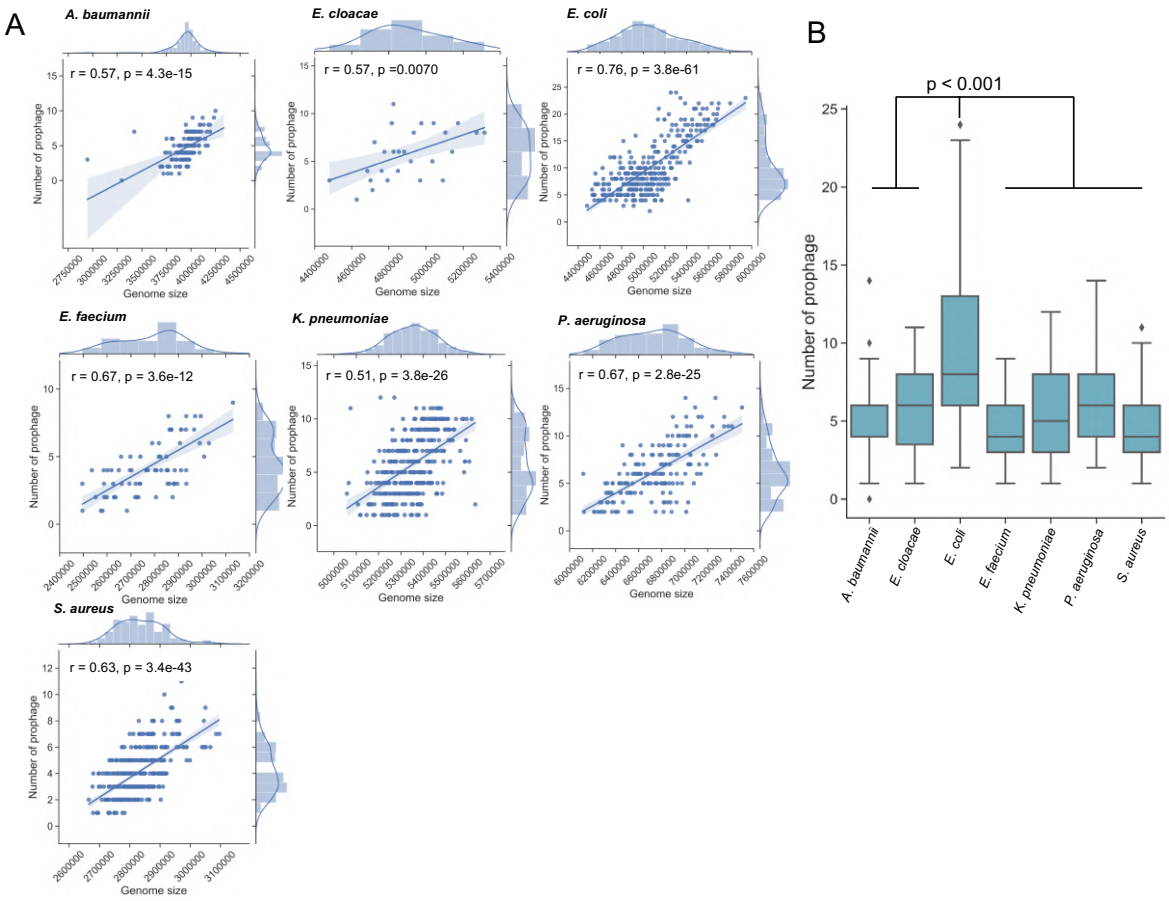

FIG 1 Comparison of genome length and the number of prophages. (A) Correlation and scatter plot of genome length excluding plasmids and the number of prophages. Plasmid samples were excluded to avoiding genome length bias. RefSeq and the complete genome were collected from the NCBI, and we evaluated the correlation between each genome length and the number of prophages obtained from PHASTER. r value indicates the Pearson's correlation coefficient. (B) Box and whisker plot comparing the number of prophages for each species. The line inner box shows the median. Welch's t-test was performed and $P < 0.05$ was considered as significant.

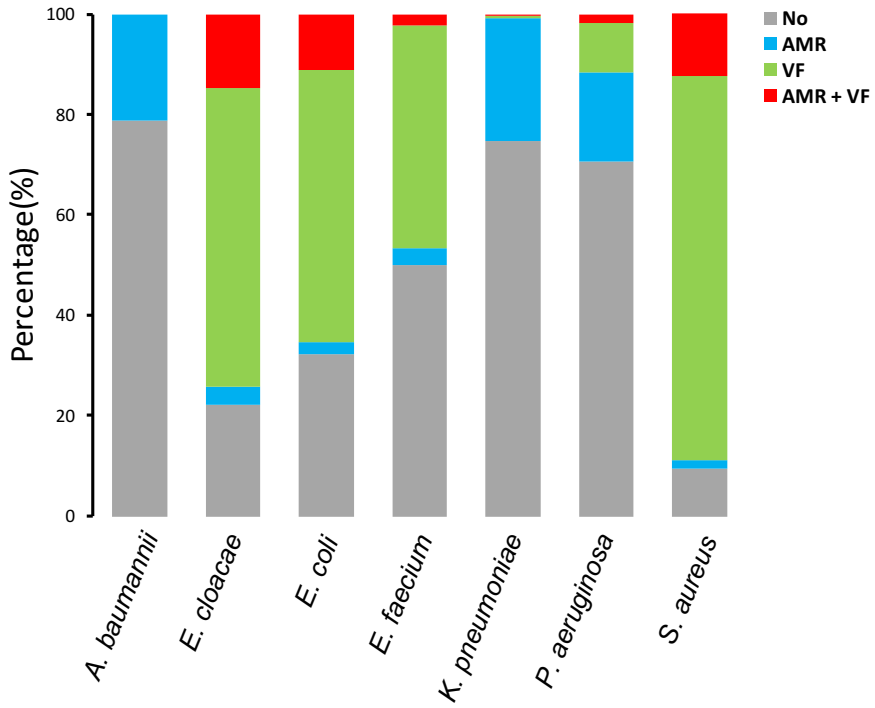

FIG 2 Proportion of the host genomes having prophages with AMR or VF genes. A bar chart showing the genome of each species that have the AMR and/or VF gene-containing prophages. Blue represents strains with AMR genes, green represents strains having VF genes, and red represents the strains having both AMR and VF genes. Gray shows strains having neither AMR nor VF genes.

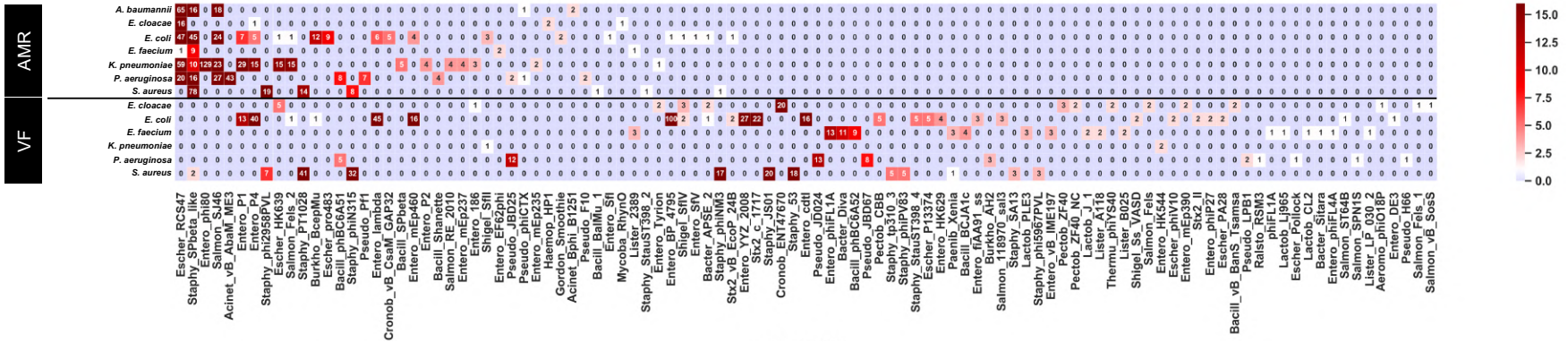

FIG 3 The type of prophages containing AMR and VF genes. Each prophage region that possessed AMR or VF genes was classified based on the most common phage in PHASTER. Heatmap shows the abundance of prophage type in each genus. The bottom represents prophage types described in PHASTER. The numeric character in each cell represents the number of detected phage names described in the most common phages on PHASTER. AMR, above the black line; VR, below the black line.

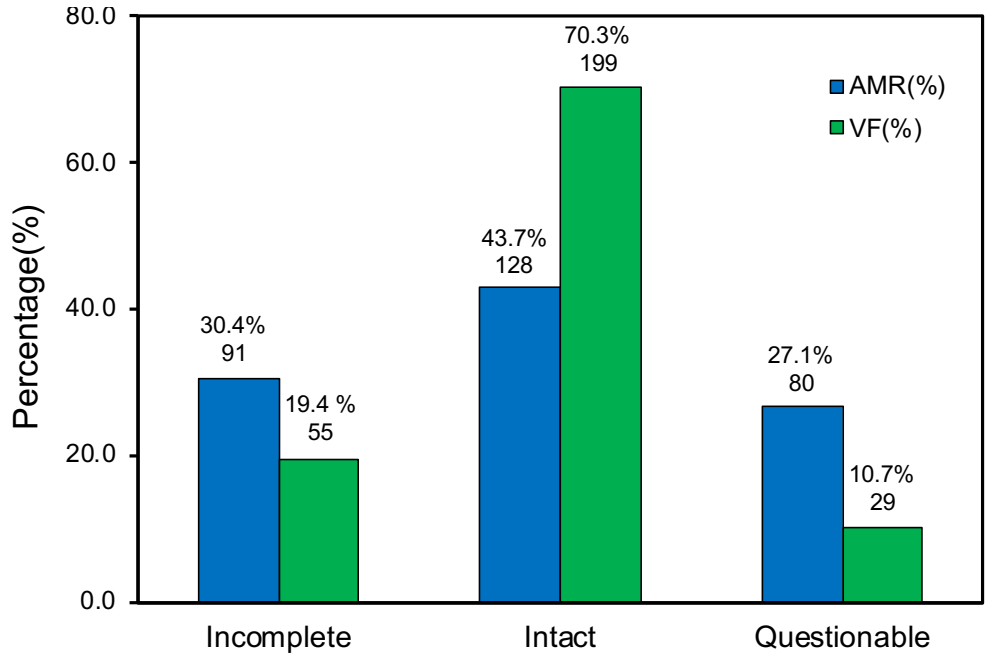

FIG 4 Analysis of prophage completeness using AMR or VF genes. Each prophage that had AMR or VF genes was classified as incomplete, intact, or questionable depending on the length and their completeness. The blue bar shows the percentage of AMR genes encoded in the prophage region, the green bar shows the percentage of prophages with a VF gene. Percentages regarded all AMR- or VF-encoding prophages as 100% and the numerical character below the percentage indicates the number of samples in each classification.

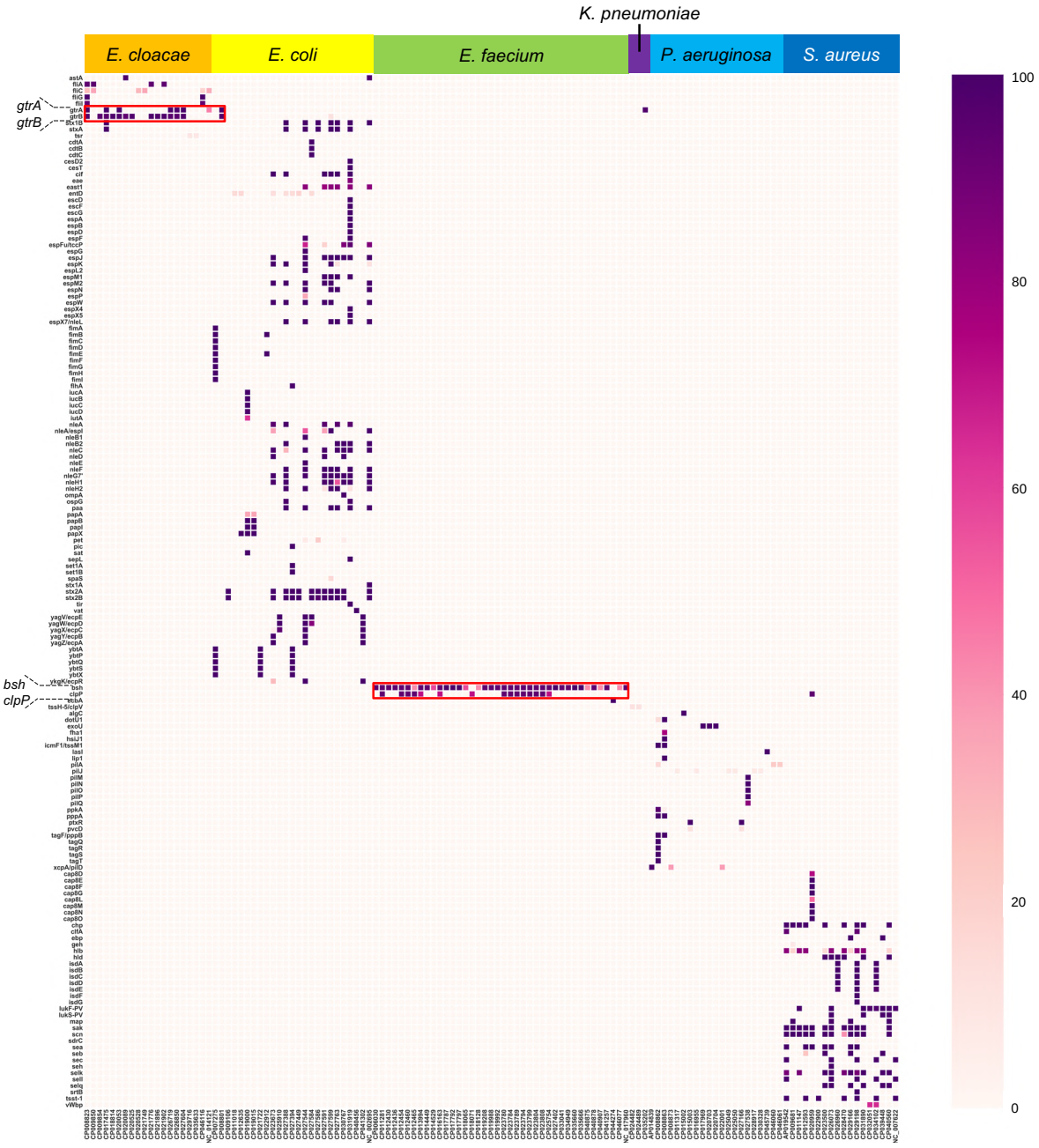

FIG 5 Distribution of VF genes in the prophage regions. DNA sequences of the prophage regions extracted from PHASTER were examined for the presence or absence of various VF genes using the VFDB database. The color density of the heatmap represents the coverage for each gene. The red frames show *gtrA*, *gtrB* and *bsh*. VF genes in *S. aureus* and *E. coli* are shown as the representative strains. All VF genes harbored by prophages are listed in Table S3.

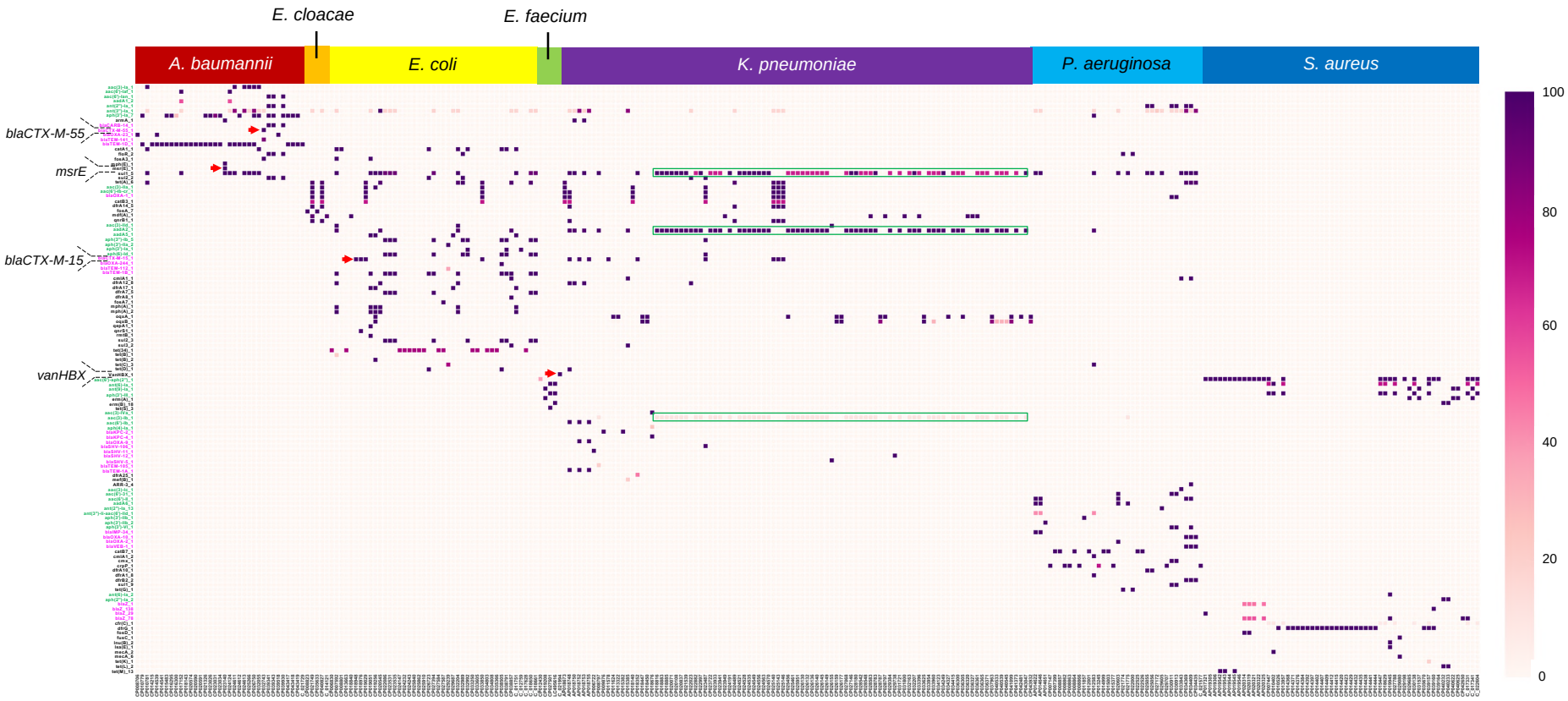

FIG 6 Distribution of AMR genes in the prophage region. We investigated the AMR genes in prophage regions extracted from PHASTER. AMR genes were detected using the ResFinder database. The color density of the heatmap represents the coverage of each gene. Green frame indicates the AMR genes cassette array detected in the *K. pneumoniae*. AMR genes, highlighted in green and magenta, show resistant genes against aminoglycoside and β-lactam, respectively. Magnified AMR gene names and red arrow represent AMR genes described in the results.

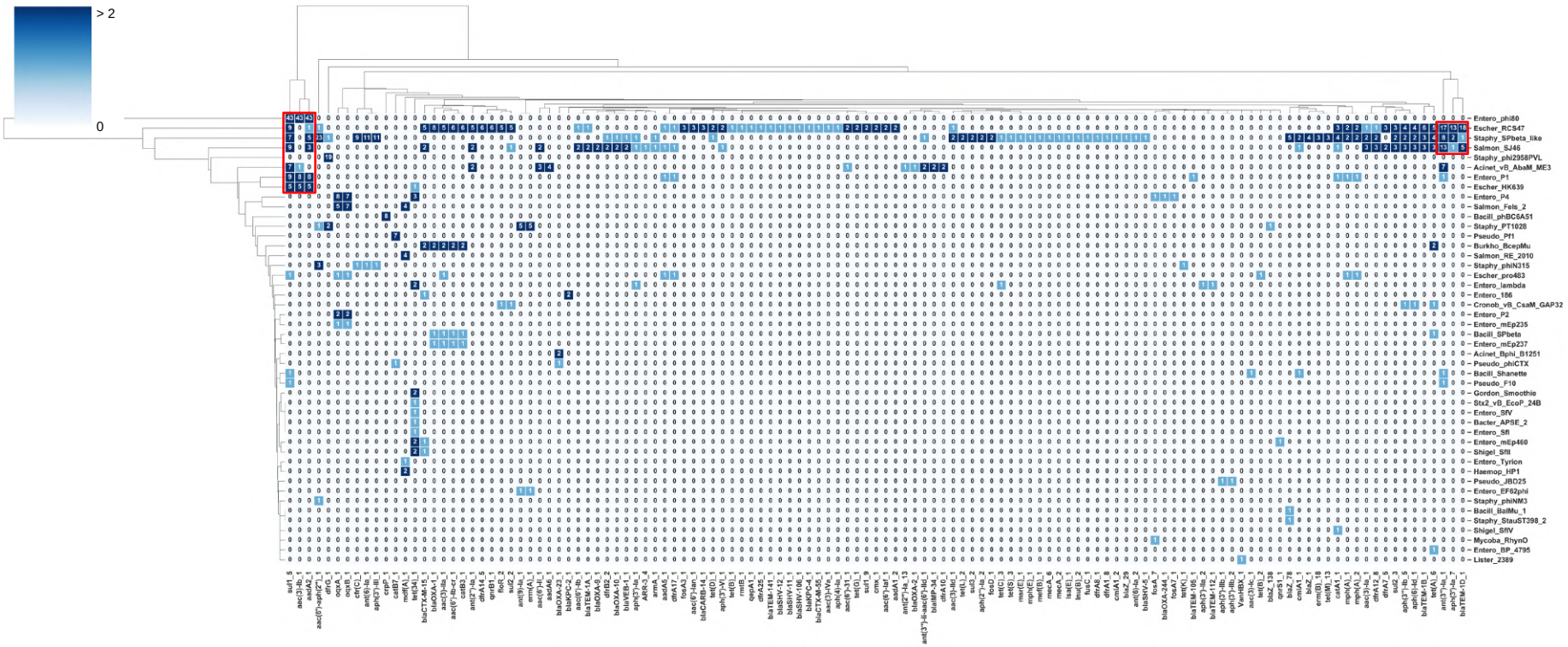

FIG 7 Cluster analysis of AMR gene-encoding prophages. Each prophage type included in all bacterial species used in this study was clustered under the same prophage name and visualized via a heatmap. The figure was created using seaborn, which is a Python module, and cluster method, which is a hierarchical clustering method (single linkage method). The numerical characters in cells mean the number of AMR genes in the indicated prophage. The red frames indicate the examples of cassette arrays of antibiotic resistance genes in the prophage region.

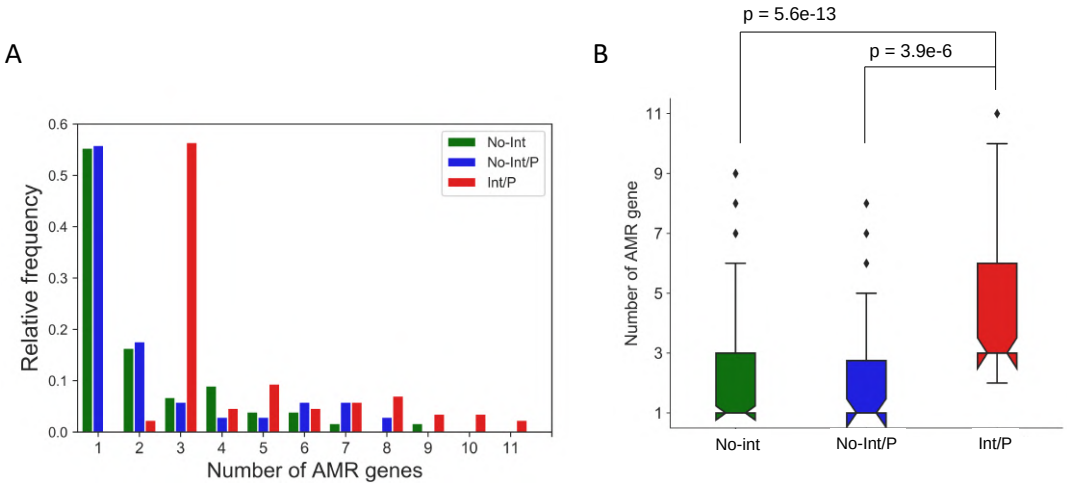

FIG 8 Comparison of the number of AMR genes with and without an integron. No-Int is a strain that does not have an integron, No-Int/P represents that the integron is located out of the prophage region, Int/P is a strain that has an integron in the prophage. Integron within a prophage was classified according to whether it has class 1 integrase or not. (A) The histogram shows the relative frequency of each number of AMR genes in each classification. (B) The boxplot shows the number of AMR genes in each classification. The notch of the boxplot represents the median, and Welch's t-test was performed with $P < 0.05$ considered as significant.

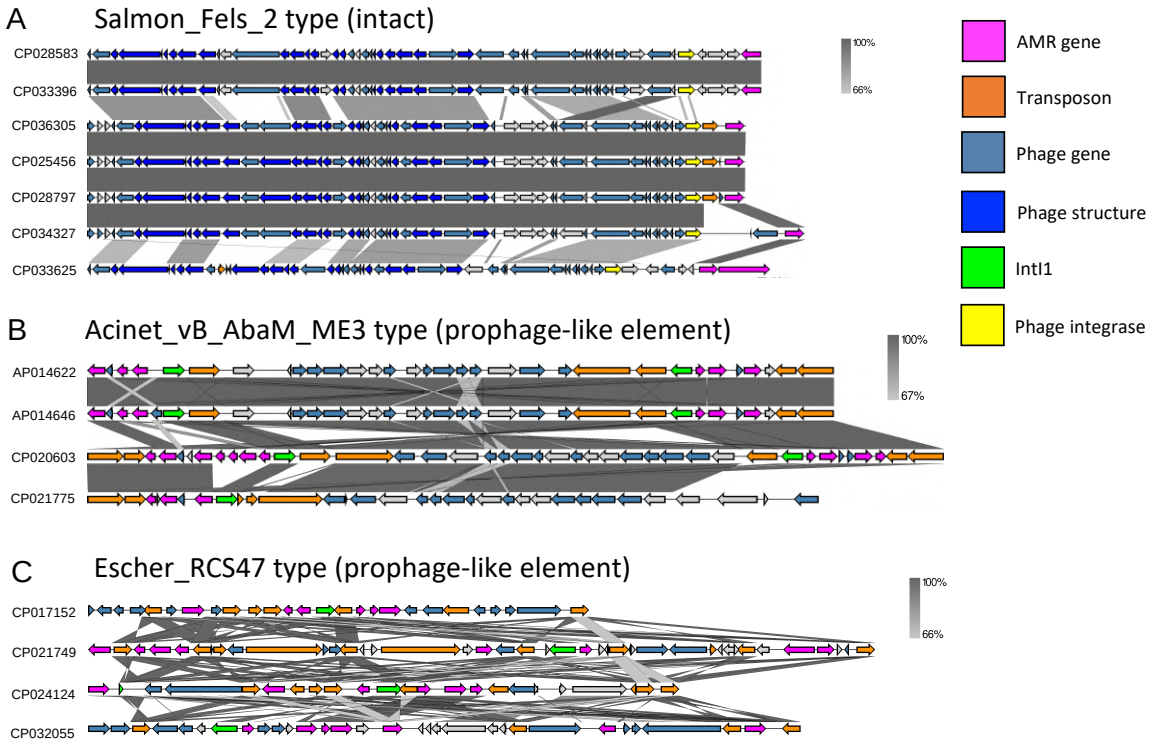

FIG 9 Comparative analysis of sequences and positional characteristics of AMR genes encoded by prophages. Representative prophage type names were selected based on genome completeness. Prophage sequences were defined via the attachment sites (*attL*, *attR*). (a) Salmon_Fels_2 prophage type was selected as an intact prophage; (b) Acinet_vB_AbaM_ME3 and (c) Escher_RCS47 were selected as prophage-like element (incomplete and questionable) prophages. The sequences of each prophage were selected at random and gray-shaded region represents the sequence similarity. IntI1 shows the class 1 integrase, which is shown by the green arrow, and phage-derived integrase is shown by the yellow arrow. AMR genes are shown in magenta. The figures were drawn using Easyfig 2.2.2.

Supplementary material

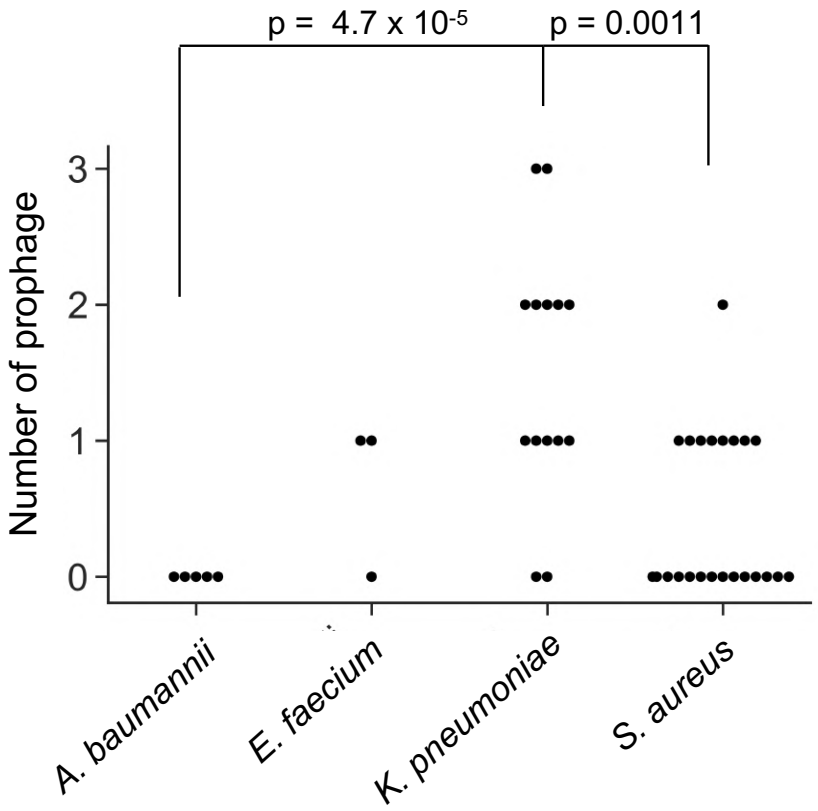

FIG S1 Comparison of the number of prophages present on the plasmid. The number of prophage regions on the plasmid was compared. Welch's t-test was performed and $P < 0.05$ was considered significant.

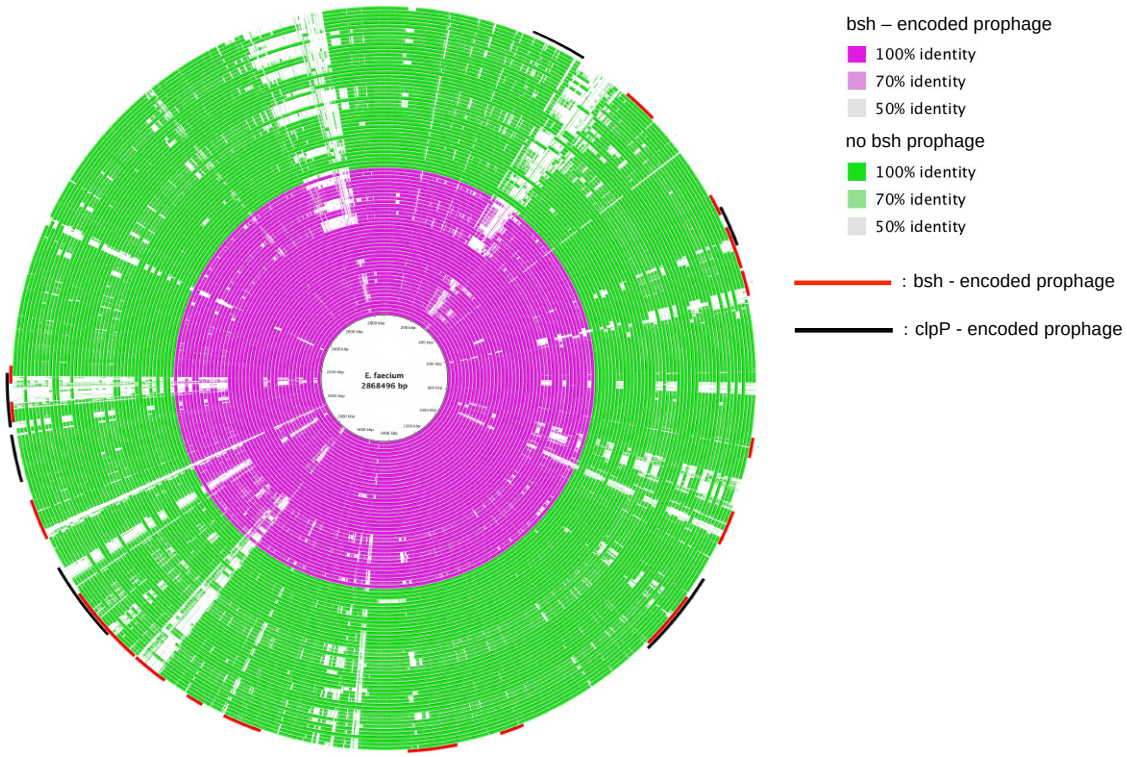

FIG S2 The mapping of prophages encoding *bsh* and *clpP* in *E. faecium*. The red circles (the inner circle) indicate the genome that includes *bsh*, and the green circles indicate the genome that does not include *bsh* in the prophage region. The red arc indicates the prophage region encoding *bsh*, and the black arc indicates that encoding *clpP*. The gap shows less homology (< 50 %) with the original genome. The figure was created using BRIG.

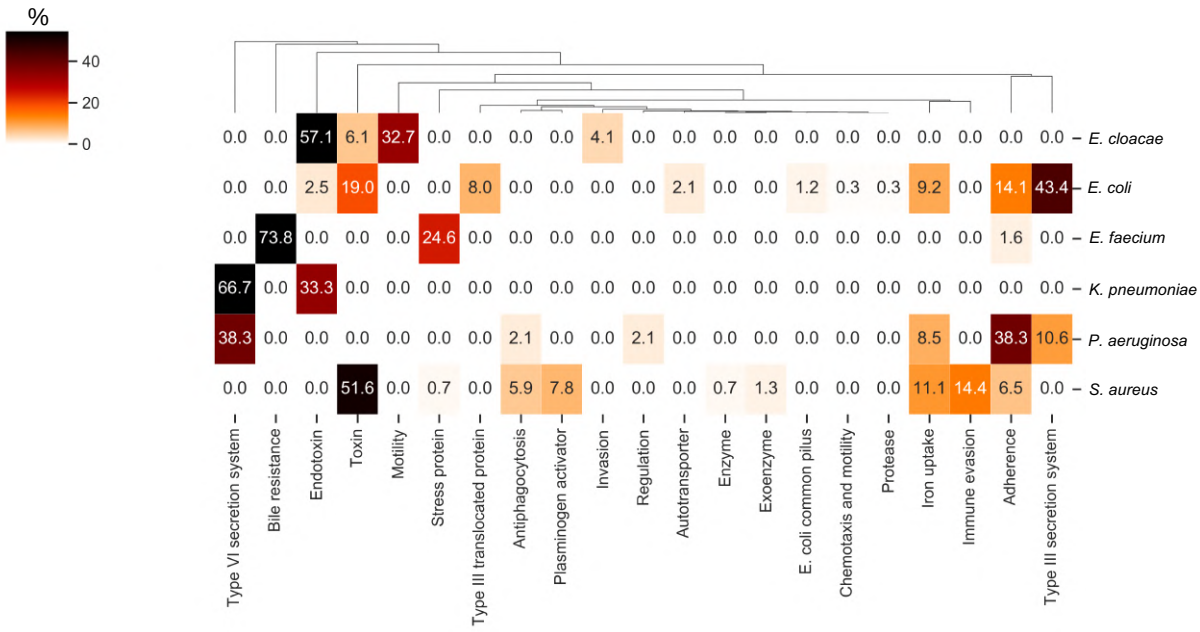

FIG S3 Classification of VF genes encoded by prophages. Each VF gene was classified based on the keywords described in the VFDB. VF genes harbored by prophages were not found in *A. baumannii*. The numerical value in the heat map indicates the percentage (%) classified into keywords for each strain.