Modular prophage interactions driven by capsule serotype select for capsule loss under phage predation

Jorge A. M. de Sousa, Amandine Buffet, Matthieu Haudiquet, Eduardo P. C. Rocha, Olaya Rendueles

HAL Id: pasteur-02991081

https://pasteur.hal.science/pasteur-02991081

Submitted on 5 Nov 2020

ISME

ARTICLE

Modular prophage interactions driven by capsule serotype select for capsule loss under phage predation

Jorge A. M. de Sousa[1] · Amandine Buffet[1] · Matthieu Haudiquet[1,2] · Eduardo P. C. Rocha[1] · Olaya Rendueles[1]

Abstract

Klebsiella species are able to colonize a wide range of environments and include worrisome nosocomial pathogens. Here, we sought to determine the abundance and infectivity of prophages of *Klebsiella* to understand how the interactions between induced prophages and bacteria affect population dynamics and evolution. We identified many prophages in the species, placing these taxa among the top 5% of the most polylysogenic bacteria. We selected 35 representative strains of the *Klebsiella pneumoniae* species complex to establish a network of induced phage–bacteria interactions. This revealed that many prophages are able to enter the lytic cycle, and subsequently kill or lysogenize closely related *Klebsiella* strains. Although 60% of the tested strains could produce phages that infect at least one other strain, the interaction network of all pairwise cross-infections is very sparse and mostly organized in modules corresponding to the strains' capsule serotypes. Accordingly, capsule mutants remain uninfected showing that the capsule is a key factor for successful infections. Surprisingly, experiments in which bacteria are predated by their own prophages result in accelerated loss of the capsule. Our results show that phage infectiousness defines interaction modules between small subsets of phages and bacteria in function of capsule serotype. This limits the role of prophages as competitive weapons because they can infect very few strains of the species complex. This should also restrict phage-driven gene flow across the species. Finally, the accelerated loss of the capsule in bacteria being predated by their own phages, suggests that phages drive serotype switch in nature.

Introduction

Phages are one of the most abundant entities on Earth. They are found in multiple environments, typically along with their host bacteria, including in the human microbiome. Many recent studies focused on virulent phages, which follow exclusively a lytic cycle. In contrast, temperate phages, which can either follow a lytic cycle, or integrate into the host genome and produce a lysogen, have been comparatively less studied. Integrated phages, hereafter referred to as prophages,

replicate vertically with the host and are typically able to protect them from infections by similar phages, the so-called resistance to superinfection [1–3]. Most prophage genes are silent and have little impact in bacterial fitness as long as there is no induction of the lytic cycle [1]. If the prophage remains in the genome for a very long period of time it may be inactivated by mutations. A few studies suggest that many prophages are inactive to some extent [4, 5]. Upon induction, some of them cannot excise (cryptic prophages), replicate, infect, or produce viable progeny. Prophage inactivation and further domestication may lead to the co-option of some phage functions by the bacterial host [6]. For instance, some bacteriocins result from the domestication of phage tails [7, 8]. In contrast, intact prophages can be induced (by either extrinsic or intrinsic factors) and resume a lytic cycle, producing viable viral particles.

Temperate phages affect the evolution of gene repertoires and bacterial population dynamics [9–11] by two key mechanisms. First, induction of prophages by a small subset of a population produces virions that can infect susceptible bacteria and thus facilitate colonization [9, 12, 13]. Second, they drive horizontal gene transfer between bacteria.

✉ Olaya Rendueles
olaya.rendueles-garcia@pasteur.fr

1 Microbial Evolutionary Genomics, Institut Pasteur, CNRS, UMR3525, Paris 75015, France

2 Ecole Doctorale FIRE—Programme Bettencourt, CRI, Paris 75005, France

Around half of the sequenced genomes contain identifiable prophages [14, 15], e.g., a third of *Escherichia coli*'s pan-genome is in prophages [16]. The frequency of prophages is higher in bacteria with larger genomes, in pathogens, and in fast-growing bacteria [15]. Lysogenization and the subsequent expression of some prophage genes may result in phenotypic changes in the host, e.g., many pathogens have virulence factors encoded in prophages [17]. Prophages may also facilitate horizontal transfer between bacteria by one of several mechanisms of transduction [1, 18, 19]. Interestingly, bacterial populations can acquire adaptive genes from their competitors by killing them with induced prophages and recovering their genes by generalized transduction [20]. While these mechanisms have been explored in many experimental and computational studies, the impact of temperate phages in the diversity of bacterial lysogens is still poorly understood.

Here, we assess the relevance of prophages in the biology of *Klebsiella* spp., a genus of bacteria capable of colonizing a large range of environments. The genus includes genetically diverse species of heterotrophic facultative aerobes that have been isolated from numerous environments, including the soil, sewage, water, plants, insects, birds and mammals [21]. *Klebsiella* spp. can cause various diseases such as urinary tract infections, acute liver abscesses, pneumonia, infectious wounds, and dental infections [22, 23]. They commonly cause severe hospital outbreaks associated with multidrug resistance, and *K. pneumoniae* is one of the six most worrisome antibiotic-resistant (ESKAPE) pathogens. The versatility of *Klebsiella* spp. is associated with a broad and diverse metabolism [24], partly acquired by horizontal gene transfer [23, 25, 26]. In addition, *Klebsiella* spp. code for an extracellular capsule that is highly variable within the species. This capsule is a high molecular weight polysaccharide made up of different repeat units of oligosaccharides. Combinations of different oligosaccharides are referred to as serotypes. In *K. pneumoniae* there are 77 serologically defined serotypes, and numbered from K1 to K77 [27] and more than 130 serotypes were identified computationally. The latter are noted from KL1 to KL130 [28, 29], and are usually referred as capsule locus types (or CLT). The capsule is considered a major virulence factor, required, for instance, in intestinal colonization [30]. It also provides resistance to the immune response and to antibiotics [31–33]. From an ecological point of view, the capsule is associated with bacteria able to colonize diverse environments [34, 35]. Its rapid diversification may thus be a major contributor to *Klebsiella*'s adaptive success, including in colonizing clinical settings.

We have recently shown that species of bacteria encoding capsular loci undergo more frequent genetic exchanges and accumulate more mobile genetic elements, including prophages [35]. This is surprising because capsules were proposed to decrease gene flow [36] and some phages are known to be blocked by the capsule [37–39]. However, several virulent phages of *Klebsiella* are known to have depolymerase activity in their tail fibers [40–44]. These depolymerases specifically digest oligosaccharidic bonds in certain capsules [45] and allow phages to access the outer membrane and infect bacteria [46]. Since depolymerases are specific to one or a few capsule types [44, 47], this implicates that some phages interact with capsules in a serotype-specific manner [42, 44, 48]. In addition, the capsule could facilitate cell infection because phages bind reversibly to it, prior to the irreversible binding to the specific cell receptor [49].

To date, the number and role of prophages in *Klebsiella*'s population biology is not well known. *Klebsiella* are interesting models to study the role of prophages, because of the interplay between the capsule, phage infections, and also the influence of the former in *Klebsiella*'s colonization of very diverse ecological niches. In this work, we sought to characterize the abundance and distribution of *Klebsiella* temperate phages, and experimentally assess their ability to re-enter the lytic cycle and lysogenize different *Klebsiella* strains. By performing more than 1200 pairwise combinations of lysates and host strains, we aim to pinpoint the drivers of prophage distribution and elucidate some of the complex interactions that shape phage–bacteria interactions in *Klebsiella*.

Results

Prophages are very abundant in the genomes of *Klebsiella*

We used PHASTER [50] to analyse 254 genomes of eight *Klebsiella* species (and two subspecies). We detected 1674 prophages, of which 55% are classified as "intact" by PHASTER and are the most likely to be complete and functional. These "intact" prophages were present in 237 out of the 254 genomes (see Methods). The remaining prophages were classed as "questionable" (20%) and "incomplete" (25%, Figs. 1a, b and S1). The complete list of bacterial genomes and prophages is available in Dataset S1. Most of the genomes were polylysogenic, encoding more than one prophage (Fig. S1). However, the number of prophages varied markedly across genomes, ranging from 1 to 16, with a median of 6 per genome. In addition, the total number of prophages in *Klebsiella* spp. varied significantly across species (Kruskal–Wallis, $df = 6$, $P < 0.001$, Fig. 1b). More specifically, the average number of "intact" prophages varied between eight for *K. oxytoca* and less than one for *K. quasipneumoniae subsp. quasipneumoniae*. *K. pneumoniae*, the most represented species of our dataset (~77% of the

Fig. 1 Prophage distribution in *Klebsiella* genus. a Rooted phylogenetic tree of *Klebsiella* species used in this study based on the core genes. **b** Average number of prophages per genome. PHASTER prediction for completeness is indicated. Numbers represent the total number of genomes analysed of each species.
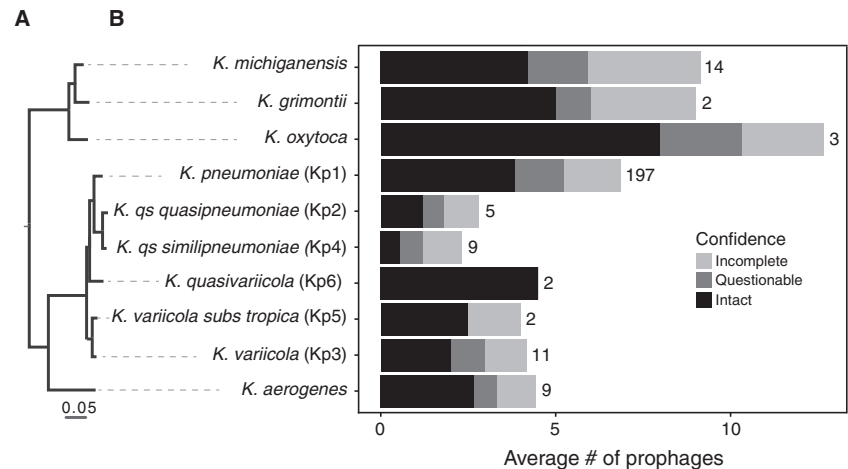
genomes), has an average of nine prophages per genome, of which four are classified as "intact" (Fig. 1b). As expected, both the number of prophages per genome and the average number of prophages per species are correlated positively with genome size (Spearman's rho = 0.49, $P < 0.001$, Spearman's rho = 0.76, $P = 0.01$, respectively) (Fig. S2A, B). When compared with the one hundred most sequenced bacterial species, the number of prophages in *Klebsiella* is very high. *K. pneumoniae* ranks within the 5th percentile of the most prophage-rich species, comparable with *E. coli* and *Yersinia enterocolitica* (Fig. S3). This shows that prophages are a sizeable fraction of the genomes of *Klebsiella* and may have an important impact in its biology.

Our experience is that prophages classed as "questionable" and "incomplete" often lack essential phage functions. Hence, all the remaining analyses were performed on "intact" prophages, unless indicated otherwise. These elements have 5% lower GC content than the rest of the chromosome (Wilcoxon test, $P < 0.001$, Fig. S2B), as typically observed for horizontally transferred genetic elements [51, 52]. Their length varies from 13 to 137 kb, for an average of 46 kb. Since temperate dsDNA phages of enterobacteria are usually larger than 30 kb [53], this suggests that a small fraction of the prophages might be incomplete (Fig. S2C). Among the "intact" prophages detected in the 35 strains analysed from our laboratory collection (Dataset S1 and Fig. S1) and isolated from different environments and representative of the genetic diversity of the *K. pneumoniae* species complex [24], we chose 11 to characterize in detail in terms of genetic architecture (Fig. 2). A manual and computational search for recombination sites (*att*) in these 11 prophages (See Methods, Fig. 2) showed that some might be larger than predicted by PHASTER. To verify the integrity of these phages, we searched for genes encoding structural functions —head, baseplate, and tail—and found them in the eleven prophages. Two of these prophages (#62 (2) and #54 (1))
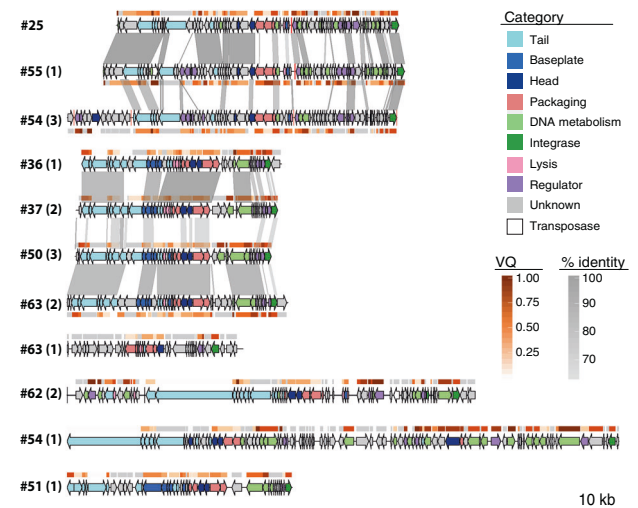

Fig. 2 Genomic organization of eleven of the prophages in this study. Numbers correspond to the host genomes, as displayed in Fig. 3. The number in parenthesis identifies the prophage in the genome. All prophages are classified as "intact" except #63 (1) ("questionable"). Genome boundaries correspond to attL/R sites. Arrows represent predicted ORFs and are oriented according to transcriptional direction. Colors indicate assigned functional categories, tRNAs are represented as red lines and the sequences are oriented based on the putative integrase localization. Local *blastn* alignments (option *dc-megablast*) are displayed between pairs of related prophages, colored according to the percentage of identity. The Viral Quotient (VQ) from pVOG is displayed below or on top of each ORF, with gray meaning that there was no match in the pVOG profiles database and thus no associated VQ value. Prophages #62 (2) and #54(1) have inserted the core gene *icd*, and the boundaries correspond to *icd* on the right and the most distal *att* site found, which is likely to be a remnant prophage border. The most proximal *att* site is also annotated (vertical black line). This figure was generated using the R package GenoPlotR v0.8.9 [103].

have a protein of *ca*. 4200 aa (Fig. 2), homologous to the tail component (gp21) of the linear phage-plasmid phiKO2 of *K. oxytoca* [54]. The prophages had integrases and were flanked by recombination sites, suggesting that they could still be able to excise from the chromosome (Fig. 2).

Nevertheless, some prophages encoded a small number of insertion sequences, which accumulate under lysogeny and degrade prophages [55].

Klebsiella spp. prophages can be released into the environment

To further characterize the prophages of *K. pneumoniae* species complex, we sought to experimentally assess their ability to excise and be released into the environment. The analysis described above identified 95 prophages in the 35 strains analyzed from our collection (Fig. 3). Among these, 40 were classed as "intact" (including the ten prophages whose genomic composition is characterized in Fig. 2). Eleven strains had no "intact" prophages, and four of these also lacked "questionable" prophages. Hence, our collection is representative of the distribution of prophages in the species, with some genomes containing one or multiple putatively functional prophages and others being prophage-free.

To test prophage induction, we grew these 35 strains and added mitomycin C (MMC) to exponentially growing cultures. Viable prophages are expected to excise and cause cell death as a result of phage replication and outburst. In agreement with PHASTER predictions, the addition of MMC to the strains lacking "intact" and "questionable" prophages showed no significant cell death (Figs. 3 and S4), with the single exception of mild cell death at high doses of MMC for strain NTUH K2044 (#56). Three of the seven strains lacking "intact" but having some "questionable" prophages showed very mild cell death upon induction, two others exhibited a dose-dependent response, and the remaining two showed rapid cell death (#44 and #35). This suggests that at least some of the "questionable" prophages are still inducible and able to kill the host. In addition, all the 24 strains with "intact" prophages showed signs of cell death *ca.* 1 h after exposure to MMC (Figs. 3 and S4). This occurred in a dose-dependent manner, consistent with prophage induction.

These results suggest that most strains have inducible prophages. To verify the release of prophage DNA to the environment, we tested the presence of 17 different phages from eleven different strains by PCR, both in induced and non-induced PEG-precipitated filtered supernatants
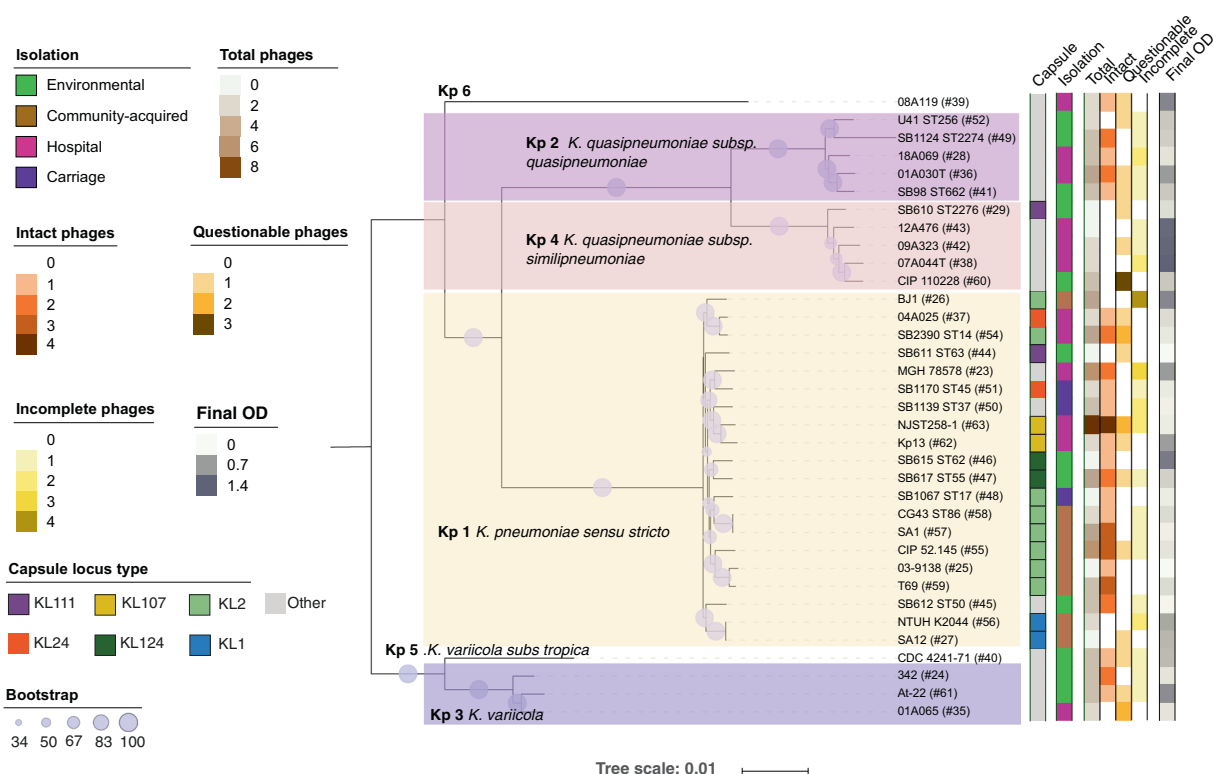

Fig. 3 Phylogenetic tree of the 35 Klebsiella strains. The tree was built using the protein sequences of 3009 families of the core genome of representative strains from the *Klebsiella pneumoniae* species complex. The first column determines the capsule locus type and the second column provides information about the environment from which it was isolated. The next columns indicate the total, "intact", "questionable", and "incomplete" prophages detected in the genomes by PHASTER. The last column shows the final absorbance of a culture after induction by mitomycin C. Background color indicates different *Klebsiella* spp. The size of the circles along the branches are proportional to bootstrap values ranging between 34 and 100. The gray color indicates other CLTs that are only present once in the dataset. The hash symbol represents the number of the strain in our collection and is used for simplicity throughout the text.

(Fig. S5A). Indeed, all 17 phages were amplified, indicating that viral genomes are released into the environment. As expected, amplification bands were weaker in the non-induced supernatant compared with the induced (Fig. S5A). We further verified the recircularization of these prophages in induced supernatants. We detected the recircularization of eight phages induced from six strains (#25, #36, #62, #63, #37, and #50), five of which were classified as "intact", and three as "questionable" by PHASTER (Fig. S5B). Despite using different primers and PCR settings, we did not obtain a clean PCR product for eight phages. This is most likely due to the aforementioned inaccuracies in the delimitation of the prophages, as these phages are detected in induced supernatants (Fig. S5A). Overall, our results suggest that most "intact" and some of the "questionable" prophages are able to induce the lytic cycle and lyse the cells.

The prophage–bacteria interaction network is sparse

Prophages may impact intraspecies competition [9]. In order to characterize this effect in the *K. pneumoniae* species complex, we produced three independent lysates of all

35 strains, by MMC induction and PEG-precipitation of the filtered supernatants (see Methods). We tested the ability of all lysates to produce a clearing on bacterial overlays of every strain. This resulted in 1225 possible lysate–bacteria combinations, with some lysates being potentially composed of multiple types of phage and different proportions across the three independent replicates. We first tested whether the number of different prophages in each strain's genome correlated with its ability to infect the other strains. To do so, we calculated the infection score of each strain, i.e., the average frequency at which its lysate infects the 35 strains of our panel. We observed a positive and significant association between the number of predicted "intact" phages and the infection score (rho = 0.16, $P < 0.001$). Overall, 21 strains out of the 35 (60%) produced a lysate that could infect another strain, but only 75 of the 1225 possible lysate–strain combinations showed clearances or plaques (Figs. 4a and S6). To confirm that plaques are caused by the activity of phages, we added lysates to growing cultures of sensitive bacteria, which resulted in bacterial death. Consistent with the action of phages, bacterial death was prevented upon the addition of citrate, a known inhibitor of phage adsorption
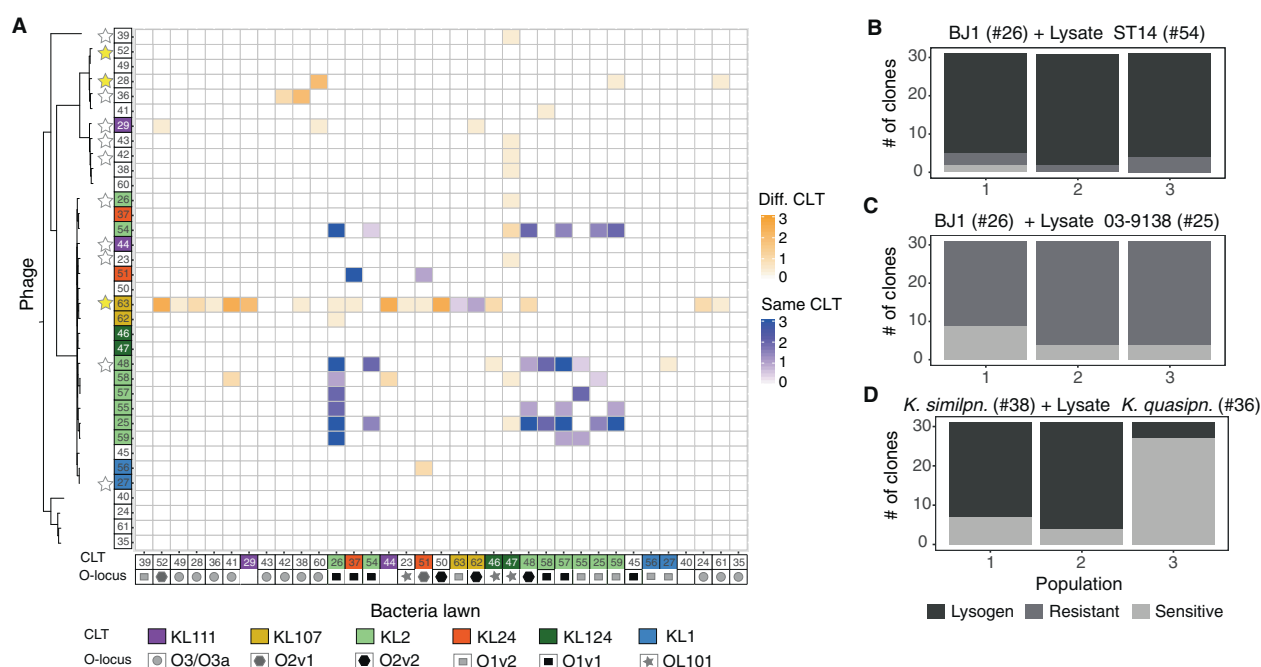

Fig. 4 Some *Klebsiella* encode viable prophages that can infect and lysogenize other strains. a Infection matrix indicating the ability of induced and PEG-precipitated supernatants of all strains to form inhibition halos on lawns of all strains. This was repeated three times with three independently produced lysates. The sum of the three experiments is shown. The strains are ordered by phylogeny, and the colors on top of numbers indicate capsule locus types (CLT). Geometric figures along the *x*-axis indicate the LPS serotype (O-locus). Stars along the *y*-axis indicate that the genome codes for putative genes involved in colicin production but lacking a lysis gene (white) or has a full colicin operon (yellow). Shades of blue indicate infections

between lysates produced by bacteria from the same CLT as the target bacteria, whereas shades of orange indicate cross-CLT infections. **b–d** Barplots indicate the number of resistant, sensitive, or lysogenized clones from each independent replicate population exposed 24 h to lysates from other strains. Each bar represents the 31 survivor clones from each of three independent experiments. Gray scale indicates whether: (i) the clones are still sensitive to the phage (light gray), (ii) they became resistant by becoming a lysogen (dark gray) as determined by cell death upon addition of MMC and PCR targeting prophage sequences, or (iii) they became resistant by other mechanisms.

(Fig. S6A). Surprisingly, 5 out of the 21 strains could be reinfected by their own lysate. This suggests the presence of ineffective repressor mechanisms protecting from super-infection, and is consistent with the observed spontaneous induction of prophages in lysogens in some of these strains (Figs. S5 and S6C).

Some of the inhibition halos observed in the experiments could result from bacteriocins induced by MMC [56]. To test if this could be the case, we searched for putative bacteriocins in the genomes and found them in 13 of the 35 genomes (Table S1, Fig. 4). However, most of them had very low (<50%) protein sequence identity with known proteins and were in genomes lacking recognizable lysis proteins. This may explain why 3 of the 13 strains failed to produce an inhibition halo on all tested strains and five showed mild inhibition of a single target strain (#47). In three of the strains (#63, #48, and #36) we could confirm phage production by recircularization (Fig. S5), showing that phages are produced. Indeed, phages from strain #36 can lysogenize strain #38 (see below, Fig. 4d), and phages from strain #48 are able to produce plaques on overlays of strain BJ1(Fig. S6B). Strain #63 has a full bacteriocin operon, with high identity to colicin E7. It is thus possible that some of the inhibition we observed from the lysates of this strain could be caused by bacter-iocins. Note that we also observed that at least two of the prophages of strain #63 are able to excise and circularize (Fig. S5 and see annotation Fig. 2). Thus, for this particular strain, we cannot precisely identify the cause of the inhibition halos. Finally, from the two remaining strains (#29 and #28, both *K. quasipneumoniae*), only #28 encodes a lysis protein, but they both mildly infect three strains each. Overall, our results suggest that for the majority of the cases shown here, inhibitions seem to be caused by phage activity and are not due to bacteriocins.

Our analysis shows that most lysates cannot infect other strains. For the 75 phage infections we do observe, only 17 (23%) occurred in all three independently produced lysates (Fig. S7), hinting to an underlying stochasticity in the induction and infection process. Interestingly, we observe that infections of target bacteria with lysates from bacteria with the same CLT seem both more reproducible and effective, com-pared with those lysates produced from bacteria with different CLT (Fig. S7), suggesting that the capsule may play a crucial role during phage infection.

Overall, our analysis shows that most lysates cannot infect other strains resulting in a sparse matrix of phage-mediated competitive interactions.

Induced *Klebsiella* prophages can lysogenize other strains

An induced prophage can, upon infection of a new host, generate new lysogens. To test the ability of the induced prophages to lysogenize other strains, we challenged *K. pneumoniae* BJ1 (#26) that lacks prophages with lysates from two strains (#54 and #25). These strains have the same CLT as BJ1 (KL2) and we have PCR evidence that they produce viral particles (see above) (Fig. S5). We grew BJ1 in contact with these two lysates, and from the surviving BJ1 cells, we isolated 93 clones from three independently challenged cultures. We reexposed these clones to a lysate from strain #54, to test if they had become resistant. Almost all clones (96%) of BJ1 could grow normally upon rechallenge (Fig. 4b), suggesting that they had acquired resistance either by lysogenization or some other mechan-ism. Most of these clones displayed significant cell death upon exposure to MMC (whereas the ancestral strain was insensitive), which is consistent with lysogenization. Ana-lyses by PCR showed that 57 out of 82 clones from the three independent BJ1 cultures exposed to #54 lysate acquired at least one of the "intact" prophages and at least four of them acquired both (shown in Fig. 2). In contrast, no BJ1 clones became lysogens when challenged with the lysate of strain #25 (as tested by exposure to MMC and PCR verification, Fig. 4c), indicating that the surviving clones became resistant to phage infections by other mechanisms. Overall, this shows that lysogenization of BJ1 is dependent on the infecting phages, with some driving the emergence of alternative mechanisms of resistance to infection in the bacterial host.

To test the taxonomic range of infections caused by induced prophages, we also exposed a Kp 4 (*K. qua-sipneumoniae subsp similipneumoniae*) (#38) to a lysate from a Kp 2 (*K. quasipneumoniae subsp quasipneumoniae*) (#36) that consistently infects #38 even if it belongs to a different subspecies and has a different capsular CLT (Fig. 4a). Survivor clones of strain #38 were lysogenized (in 61 out of 93 surviving clones) exclusively by one of the phages present in strain #36, as confirmed by PCR of the phage genome (Fig. 4d). Interestingly, four of these lyso-genized clones did not display cell death when exposed to MMC, suggesting that their prophages might not be fully functional. Taken together, our results show that some *Klebsiella* prophages can be transferred to and lysogenize other strains, including from other subspecies.

Resistance to superinfection and bacterial defenses do not explain the interaction matrix

To investigate the determinants of infections in the inter-action matrix, we first hypothesized that resident prophages could hamper infections by lysates of other strains. Con-sistently, the most sensitive strain in our panel (#26) does not have any detectable prophages, rendering it sensitive to infection by numerous lysates. However, we found no negative association between the number of prophages in

the target strain and the number of times it was infected (rho = 0.04, P = 0.166). Repression of infecting phages is expected to be highest when there are similar resident prophages in the strain. Even if closer strains are more likely to carry the same prophages, the interaction matrix clearly shows that phages tend to infect the most closely related strains. To determine if the presence of similar prophages shapes the infection matrix, we calculated the genetic similarity between all "intact" prophages using weighted gene repertoire relatedness (wGRR) (see Methods). The frequency of infection was found to be independent from the similarity between prophages (determined as higher than 50% wGRR for phage pairs, Odds ratio = 0.55, P value = 0.055) when analysing only "intact" prophages or also including the "questionable" (Fig. S8A). Ultimately, resident phages may repress incoming phages if they have very similar repressors, but pairs of dissimilar phages (wGRR < 50%) with similar repressors (>80% sequence identity) are only ~0.3% of the total (see Methods, Fig. S8B). Thus, the analysis of both phage sequence similarity and similarity between their repressors suggests that the observed interaction matrix is not strongly influenced by the resistance to superinfection provided by resident prophages.

Bacterial defense systems block phage infections and are thus expected to influence the interaction matrix. To study their effect, we analysed systems of CRISPR-Cas and restriction-modification (R-M), since these are the most common, the best characterized to date, and those for which tools for their computational detection are available [57, 58]. We found CRISPR-Cas systems in 8 of the 35 strains and tested if the strains that were infected by the lysates of other bacteria were less likely to encode these systems. We observed no correlation between the number of unsuccessful infections and the presence of CRISPR-Cas systems in a genome (P > 0.05, Wilcoxon test). Accordingly, the majority of strains (77%) lack spacers against any of the prophages of all the other strains and we found no correlation between the presence of CRISPR spacers targeting incoming phages and the outcome of the interaction (i.e., infection or not) (Odds ratio = 1.28 Fisher's exact test, P value = 1, Dataset S1, Fig. S9A). Actually, only 5% of the pairs with null interactions (no infection) concerned a target strain with CRISPR spacers matching a prophage in the lysate-producing strain. This indicates that CRISPR-Cas systems do not drive the patterns observed in the infection matrix.

The analysis of R-M systems is more complex because their specificity is often harder to predict. We searched for these systems in the 35 genomes and identified the following systems: 37 type I, 158 type II, 4 type IIG, 7 type III, and 7 type IV. We then investigated whether the systems present in the target strains could protect from phages in the lysates. Almost all the strains studied here have systems that

could target at least one prophage from the lysate-producing strains (Fig. S9B, see Methods). However, the joint distribution of R-M systems and targeted prophages does not explain the outcome observed in the infection matrix (Odds ratio = 0.78, Fisher's exact test on a contingency table, P value = 0.4, Fig. S9). We thus conclude that the distribution of either CRISPR-Cas or R-M bacterial defense systems in the bacterial strains targeted by the phages in the lysates does not explain the network of phage–bacteria interactions we observe.

The capsule plays a major role in shaping phage infections

The first bacterial structures that interact with phages are the capsule and the LPS. We thus tested whether their serotypes shape the infection network of the *Klebsiella* prophages. We used Kaptive to serotype the strains from the genome sequence and tested if the infections were more frequent when the target bacteria and the one producing the lysate were from the same serotype. We found no significant effect for the LPS serotypes (Fig. 4a, Fisher's exact test P = 0.37). In contrast, we observed 35 cross-strain infections between lysates and sensitive bacteria from the same CLT out of 105 possible combinations (33%), whereas only 3.6% of the possible 1120 inter-CLT infections were observed (40 infections in total, Fig. 4a, P < 0.0001, Fisher's exact test). For example, strain #37 was only infected by lysates produced by the other strain with the KL24 CLT (#51) (Fig. 2). Similarly, strains #57, #55, and #25 were only infected by lysates from strains of the same CLT (KL2) (Fig. 4a). These results are independent of the genetic relatedness, as we observed infections between strains with the same CLT that are phylogenetically distant. Intriguingly, the lysate of one strain alone (#63) produced an inhibition halo in lawns of fifteen strains from different CLTs, but (as mentioned above) this may be caused by a bacteriocin present in this strain's genome. To control for the putative effect of bacteriocins in the association between infections and the CLT, we restricted the analysis to strains lacking bacteriocin homologs (N = 22) and those with a CLT determined with high confidence by Kaptive (N = 29) (see Methods). We confirmed in both cases that CLT is positively associated with cross-strain infections (P < 0.0001, Fisher's exact test for both).

The specificity of induced prophages for strains with a CLT similar to the original cell could be explained by the presence of depolymerases in their tail fibers. Visual observation of plaques showed that the latter are small and are not surrounded by enlarged halos typically indicative of depolymerase activity (Fig. S6). We also searched the genomes of the prophages for published depolymerase domains using protein profiles and a database of known depolymerases (see Methods, Table S2). We only found hits to two protein
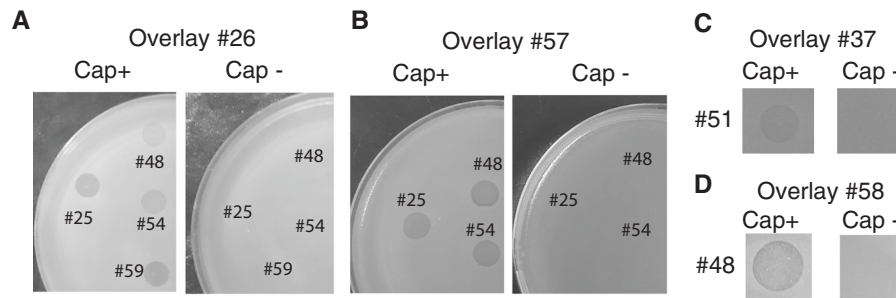
Fig. 5 The capsule is required for phage infection. Lawns of different strains, wild type (Cap+) and capsule null mutant (Cap-) in contact with PEG-precipitated supernatant of other strains. Isogenic capsule mutants of strains #26 and #37 were constructed by an in-frame deletion of *wza* gene, and a *wcaJ* deletion for strains #57 and #58 (see Methods).

profiles (Pectate_lyase_3 and Peptidase_S74) with poor *e*-value (>10^{-10}) and a low identity homolog (<50%) to one protein sequence (YP_009226010.1) (Tables S3 and S4). Moreover, these homologs of depolymerases were found in multiple prophages that are hosted in strains differing in CLTs, suggesting that they are either not functional, or that they do not target a specific CLT. It is also possible that the CLT has switched (i.e., changed by recombination [59]) since the phage originally infected the strain. This suggests that the depolymerases present in these prophages (or at least those that we could identify) do not explain the patterns observed in the interaction matrix, raising the possibility that novel depolymerases remain to be found.

It is often mentioned in the literature that capsules protect against phages [39]. However, the results above show that induced prophages tend to infect strains of similar CLT, and capsule mutants of the strain Kp36 were previously shown to become resistant to the virulent phage 117 [60]. Hence, we wondered if strains lacking a capsule would also be immune to temperate phages, and so we tested whether isogenic capsule mutants from different CLTs (KL2, KL30, and KL1) were sensitive to phages present in the lysates. We verified by microscopy and biochemical capsule quantification that these mutants lacked a capsule and then challenged them with the lysates of all 35 strains. Wild type strains #56 (K1) and #24 (K30) are resistant to all lysates, as are their capsule mutants. Capsule mutants from #26 (BJ1), #57 and #58 (all KL2), and #37 (KL24) were resistant to phages, whilst their respective wild-types are all sensitive to phage infections (Fig. 5). Hence, phages targeting these strains only infect when a capsule is present. These results show that the loss of the capsule does not make bacteria more sensitive to phages. On the contrary, the capsule seems required for infection.

Phage predation stimulates capsule loss

Given that non-capsulated mutants are resistant to phages (as shown above and also observed in other studies [60]),

we hypothesized that phages that infect a particular strain could drive the loss of its capsule. To test this, we performed a short-term evolution experiment (ten days, ca 70 generations). We assessed the emergence of non-capsulated bacteria in three different strains: (i) a strain with no temperate phages (BJ1, #26); (ii) a strain that produces phages infecting many strains including itself (#54) even in the absence of MMC induction (Fig. S6C); and (iii) a strain that is resistant to its own phages (#36). Both strains #36 and #54 produce infectious phages which can lysogenize other strains (#38 and #26, respectively, Fig. 4b). Six independent populations of each strain were evolved in four different environments: (i) LB, (ii) LB supplemented with 0.2% citrate to inhibit phage adsorption, (iii) LB with MMC to increase the phage titer, and (iv) LB with MMC and 0.2% citrate to control for the effect of faster population turnover due to prophage induction and the subsequent cellular death. We expected that passages in rich medium might lead to non-capsulated mutants, even in the absence of phages, as previously observed [61–63]. This process should be accelerated if phage predation drives selection of the non-capsulated mutants (presence of phages, absence of citrate, in LB). It should be further accelerated if the intensity of phage predation increases (under MMC). As expected, all strains progressively lost their capsule albeit at different rates (Fig. 6a). To allow comparisons between treatments and strains, we calculated the area under the curve during the first five days, where most of the capsule loss took place (Fig. 6b). Strain BJ1 lacks prophages and shows no significant differences between the treatments. Strain #54 lost its capsule faster when prophages were induced (MMC) and citrate relieved this effect in accordance with the hypothesis that the speed of capsule loss depends on the efficiency of phage infection (MM+citrate). Similarly, in the treatment with citrate the capsule is lost at a slower rate than in the LB treatment. In the latter, the few events of spontaneous prophage induction generate a basal level of predation that is sufficient to increase the rate of loss of the capsule (albeit not to the levels of the MMC treatment). Note that the

Fig. 6 Loss of capsule in three Klebsiella strains. a Ratio of capsulated clones throughout the ten days before daily passages of each culture. Shades of green represent the different environments in which evolution took place. MMC stands for mitomycin C. Full lines represent the average of the independent populations of the same strain and environment (shades of green). Dashed lines represent each of the independent populations. **b** Area under the curve during the first five days of the experiment.
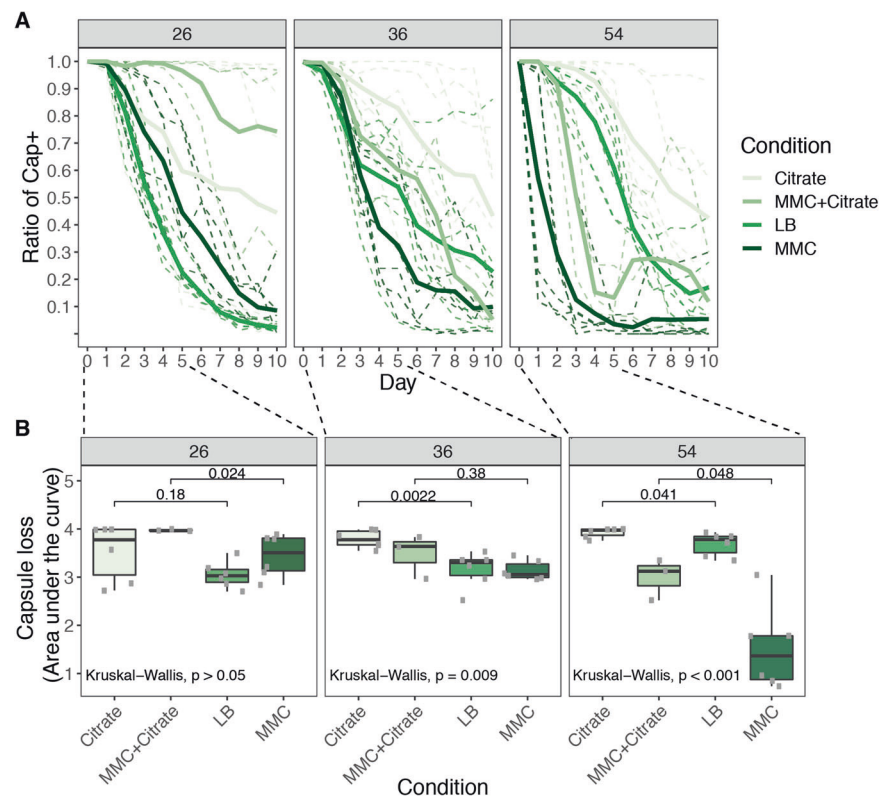
combination of citrate and MMC did not significantly affect bacterial growth, and thus the effect we observe seems to be caused by phage predation (Figs. S6 and S10). Finally, strain #36 showed no significant difference between the experiments with MMC and LB. This suggests that the amount of phages in the environment does not affect the rate of capsule loss in this strain, consistent with it being insensitive to its own phage. Intriguingly, adding citrate lowered the rate of capsule loss in this strain, a result that suggests that even if phage infection is inefficient, there may be small deleterious impacts due to the presence of phages. Taken together, these results show that effective predation by induced prophages selects for the loss of the capsule in the lysogen.

Discussion

We provide here a first comprehensive analysis of the distribution of prophages in the *Klebsiella* genus, their genetic composition and their potential to excise, infect and lysogenize other strains. The number of prophages varies significantly across the species of the genus, but most genomes of *Klebsiella* encode for a phage or its remnants. *K. pneumoniae* is one of the species with more prophages among widely sequenced bacteria, suggesting that temperate phages are particularly important for its biology. This rapid turnover of

prophages, already observed in other species, may contribute to the phenotypic differences between strains. Since many of these prophages seem to retain the ability to excise, form viable virions, and lysogenize other bacteria, they could spur adaptation when transferring adaptive traits (including antibiotic resistance in clinical settings). Further work will be needed to assess the phenotypic consequences of lysogeny and the frequency of transduction, but these results already show that the contribution of prophages to *Klebsiella* genomes is significant.

Phage–bacteria interactions shape a myriad of biological processes. Several recent studies have detailed infection networks to understand the ecological traits and the molecular mechanisms shaping them [64–66]. These analyses use isolated phages from laboratory stocks, or phages recovered from coevolution experiments [64]. In addition, these studies tend to focus on virulent phages rather than temperate, either because they envision some sort of phage therapy or because virulent phages give simpler phenotypes. Several recent computational studies have described the different phage families and the beneficial traits they may impart to their hosts, such as virulence factors [16, 65, 67, 68]. A few have explored the natural diversity of temperate phages of a species and experimentally assessed their ability to cross-infect other strains [65]. Here, we sought to generate a network of temperate phage-mediated bacterial interactions in naturally infected

Klebsiella spp, including competitive killing through phage induction and also the transfer of prophages between strains. These results are especially pertinent for *K. pneumoniae* in clinical settings because many antibiotics stimulate prophage induction [17, 69, 70] and facilitate phage infection [71, 72]. Also, phages in the mammalian gut, the most frequent habitat of *K. pneumoniae*, tend to be temperate and result from in situ prophage induction [65, 73].

We studied a large cohort of strains from the *K. pneumoniae* species complex, representing considerable diversity in terms of the number and types of prophages present in lysogenic strains. However, the extrapolation of some of the results presented here to the *Klebsiella* genus as a whole must be done with care. Further work, and a larger picture, will be possible once genomes from other clades of the genus are available. Moreover, since it would be unfeasible to generate mutants for all the strains, or perform evolution experiments with the entire cohort, some experimental assays performed here were limited to a reduced number of strains and environments. It is thus possible that some observations may not be consistently obtained with other genetic backgrounds, or ecological contexts (e.g., spatially structured environments). Nevertheless, our work considers a total of 1225 lysate–bacteria interactions, which already allows the characterization of a wide range of possible outcomes.

Our experimental setup was specifically designed to address the natural variation in the ability of lysates to infect other strains. In order to capture this variation, the network of infections was performed thrice with three independently generated lysates, which led to stochastic variation in the infection outcomes. This has been rarely characterized before for poly-lysogens. One possible explanation for the observed stochasticity in the outcomes of infections could be that the resident phages are unfit due to the accumulation of deleterious mutations during lysogeny. This could affect, for instance, the number of produced virions, and thus the infection efficiency. Alternatively, there could be a noteworthy degree of natural variation in the frequencies of each induced phage across the independently generated lysates. This means that the underrepresentation of certain phages could lead to unsuccessful infections. Having multiple prophages is expected to be costly (in terms of gene expression and cell death by induction), but extends the range of competitors that can be affected by prophage induction. Finally, infections of target bacteria with lysates from bacteria with the same CLT seem both more reproducible and effective, compared with those lysates produced from bacteria with different CLT (Fig. S7), suggesting that stochasticity in infections can also be a consequence of capsule–phage interactions, e.g., if the former is unevenly distributed throughout the cell envelope [74] or if there is stochasticity in its expression [75, 76]. Future experiments tackling diverse ecological scenarios (e.g., cocultures between three or more strains) will help understanding the causes and consequences of multiple infections producing poly-lysogens for competitive and evolutionary interactions between strains.

Overall, we found that 60% of the strains could produce at least one lysate that led to phage infection of at least one other strain. The number of strains able to produce phages is comparable with what is observed in other species like *Pseudomonas aeruginosa* (66%) [77] and *Salmonella enterica* (68%) [78]. However, we only observed 6% of all the possible cross-infections. This is much less than in *P. aeruginosa*, where a set of different lysogenic strains derived from PA14 was infected by *ca* 50% of the lysates tested [77]. Similarly, in *S. enterica*, ~35% of cross-infections were effective [78]. This suggests that the likelihood that a prophage from one strain is able to infect another *Klebsiella* strain is relatively small. Hence, when two different *Klebsiella* strains meet, they will be very often immune to the prophages of the other strains. This implicates that prophages would be less efficient in increasing the competitive ability of a *Klebsiella* strain than in other species. Finally, this also implies that phage-mediated HGT in *Klebsiella* may not be very efficient in spreading traits across the species, which means that, in some situations, the capsule could slow down evolution by phage-mediated horizontal gene transfer.

Interestingly, 5 out of 35 strains can be infected by their own induced prophages. It is commonly assumed that lysogens are always resistant to reinfection by the same phage [79], but the opposite is not unheard of. It has been previously reported that *E. coli* strains could be reinfected by the same phage twice, both by phage lambda [80] and by phage P1 [81]. At this time, we can only offer some speculations for this lack of superinfection immunity. For example, it has been estimated that there are only ten free dimmers of Lambda's cI repressor in a typical cell [82]. The infection of several phages at the same time may titer this limited amount of repressor, such that some incoming phages are able to induce resident prophages to enter a lytic cycle. This could be amplified when prophages accumulate deleterious mutations in the repressor or in the binding sites of the repressor [83–85], further decreasing its ability to silence the prophage or the incoming phage. This could also explain why we observe frequent prophage spontaneous induction (strain #54, Fig. S6C).

The small number of cross-infectivity events could not be attributed to bacterial defense systems (R-M or CRISPR-Cas) or LPS serotypes. The similarity between prophages or their repressor proteins, potentially facilitating superinfection immunity [77], also failed to explain the observed infection patterns. Instead, the relatively few cases of phage infection in the matrices are grouped in modules that seem determined by the capsule composition, since there is a vast overrepresentation of infections between bacteria of the same CLT. This seems independent of their genetic

relatedness, since we observed infections between strains with the same CLT that are phylogenetically distant. This CLT specificity is consistent with a recent study that focuses on a carbapenemase-producing *K. pneumoniae* CG258, in which very few phages could lyse bacteria with different CLT [86]. The CLT specificity of phage-encoded depolymerases [40–42, 44, 47, 48, 87, 88] could explain these results. Yet, in our set, few of such enzymes were detected, they exhibited very low identity to known depolymerases, and did not seem to correlate with the CLT. At this stage, it is difficult to know if depolymerases are rare in temperate phages or if they are just too different from known depolymerases, since these were mostly identified from virulent phages. If depolymerases are indeed rare in temperate phages, this raises the important question of the alternative mechanisms that underlie their capsule specificity.

Most of the literature on other species concurs that the capsule is a barrier to phages by limiting their adsorption and access to cell receptors [37–39]. In contrast, our results show that the capsule is often required for infection by *Klebsiella* induced prophages. This is in agreement with a recent study where inactivation of *wcaJ*, a gene essential for capsule synthesis, rendered the strain resistant to phage infection by a virulent phage [60]. Specific interactions between the phage and the capsule could be caused by the latter stabilizing viral adsorption or allowing more time for efficient DNA injection [49]. This may have resulted in phages selecting for the ability to recognize a given capsule serotype. The specificity of the interactions between temperate phages and bacteria, caused by capsule composition, has outstanding implications for the ecology of these phages because it severely limits their host range. Indeed, we report very few cases (*ca.* 3%) of phages infecting strains with different CLTs, and this could be an overestimation if we discard strain #63 because of its potential bacteriocin activity. The consequence of this evolutionary process is that phage pressure results in selection for the loss of capsule because non-capsulated bacteria are resistant to phage infection. Our experimental evolution captures the first steps of this coevolution dynamic, suggesting that phage predation selects very strongly for capsule loss in the infected strains. Since capsules are prevalent in *Klebsiella*, this suggests that it may be reacquired later on.

Our results might also provide insights regarding the possible use of phages to fight the increasing challenge of antibiotic resistance in *Klebsiella* infections. Although more work is needed to understand how to best use virulent phages to control *Klebsiella* infections, our results already hint that phage therapy may, at least in a first step, lead to capsule loss. While such treatments may be ineffective in fully clearing *Klebsiella* infections (due to the large diversity of existing serotypes), they can select for

non-capsulated mutants. The latter are expected to be less virulent [89], because the capsule is a major virulence factor in *K. pneumoniae*. The increase in frequency of non-capsulated mutants may also increase the efficiency of traditional antimicrobial therapies, as the capsule is known to increase tolerance to chemical aggressions, including antibiotics and cationic antimicrobial peptides produced by the host [89].

Materials and methods

Strains and growth conditions

We used 35 *Klebsiella* strains that were selected based on MLST data, and representative of the phylogenetic and clonal diversity of the *K. pneumoniae* species complex [24]. Strains were grown in LB at 37° and under shaking conditions (250 rpm).

Genomes

254 genomes of *Klebsiella* species (of which 197 of *K. pneumoniae*) and one *Cedecea* sp (outgroup) were analysed in this study. This included all complete genomes belonging to *Klebsiella* species from NCBI, downloaded February 2018, and 29 of our own collection [24]. We corrected erroneous NCBI species annotations using Kleborate typing [28]. 253 genomes were correctly assigned to its species, with a "strong" level of confidence, as annotated by Kaptive. Only two genomes were classified with a "weak" level of confidence by the software. All information about these genomes is presented in Dataset S1.

Identification of prophages

To identify prophages, we used a freely available computational tool, PHASTER [50], and analysed the genomes on September 2018. The completeness or potential viability of identified prophages are identified by PHASTER as "intact", "questionable" or "incomplete" prophages. All results presented here were performed on the "intact" prophages, unless stated otherwise. Results for all prophages ("questionable", "incomplete") are presented in the supplemental material. Primers used for phage detection and recircularization are presented in Table S5.

Prophage characterization

(i) *Prophage delimitation*. Prophages are delimited by the *attL* and *attR* recombination sites used for phage recircularization and theta replication. In some instances, these sites were predicted by PHASTER. When none were found,

we manually searched for them either by looking for similar *att* sites in related prophages or by searching for interrupted core bacterial genes. (ii) *Functional annotation* was performed by combining multiple tools: *prokka* v1.14.0 [90], pVOG profiles [91] searched for using *HMMER* v3.2.1 [92], the PHASTER Prophage/Virus DB (version Aug 14, 2019), *BLAST 2.9.0+* [93], and the *ISFinder* webtool [94]. For each protein and annotation tool, all significant matches (*e*-value < 10^{-5}) were kept and categorized in dictionaries. If a protein was annotated as "tail" in the description of a matching pVOG profile or PHASTER DB protein, the gene was categorized as tail. Results were manually curated for discrepancies and ties. For proteins matching more than one pVOG profile, we attributed the Viral Quotient (VQ) associated to the best hit (lowest *e*-value). The VQ is a measure of how frequent a gene family is present in phages, and ranges from zero to one with higher values meaning that the family is mostly found in viruses. (iii) *Repressor identification*. Repressors in the prophages were detected using specific HMM protein profiles for the repressor, available in the pVOG database [91]. We selected those with a VQ higher than 0.8. These profiles were matched to the *Klebsiella* intact prophage sequences using HMMER v3.1b2, and we discarded the resulting matches whose best domain had an *e*-value of more than 0.0001 or a coverage of less than 60%. We further selected those with at least one of the following terms in the descriptions of all the proteins that compose the HMM: immunity, superinfection, repressor, exclusion. This resulted in a set of 28 profiles (Table S6). The similarity between repressors was inferred from their alignments (assessed with the *align.globalxx* function in from the *pairwise2* module in biopython, v1.74), by dividing the number of positions matched by the size of the smallest sequence (for each pair of sequences).

Core genome

(i) *Klebsiella* spp. (*N* = 255) (Figs. 1a and S1). The core genome was inferred as described in [95]. Briefly, we identified a preliminary list of orthologs between pairs of genomes as the list of reciprocal best hits using end-gap free global alignment, between the proteome of a pivot and each of the other strains proteome. Hits with less than 80% similarity in amino acid sequences or more than 20% difference in protein length were discarded. (ii) *Laboratory collection of Klebsiella genomes* (*N* = 35) (Fig. 4). The pangenome was built by clustering all protein sequences with Mmseqs2 (v1-c7a89) [96] with the following arguments: *-cluster-mode 1* (connected components algorithm), *-c 0.8 –cov-mode 0* (default, coverage of query and target >80%) and *–min-seq-id 0.8* (minimum identity between sequences of 80%). The core genome was taken from the pangenome by retrieving families that were present in all genomes in single copy.

Phylogenetic trees

To compute both phylogenetic trees, we used a concatenate of the multiple alignments of the core genes aligned with MAFFT v7.305b (using default settings). The tree was computed with IQ-Tree v1.4.2 under the GTR model and a gamma correction (GAMMA). We performed 1000 ultrafast bootstrap experiments (options *–bb* 1000 and *–wbtl*) on the concatenated alignments to assess the robustness of the tree topology. The vast majority of nodes were supported with bootstrap values higher than 90% (Fig. S1). The *Klebsiella* spp. (*N* = 255) tree had 1,106,022 sites, of which 263,225 parsimony-informative. The phylogenetic tree of the laboratory collection strains (*N* = 35) was built using 2,800,176 sites, of which 286,156 were parsimony-informative.

Capsule serotyping

We used Kaptive, integrated in Kleborate, with default options [28]. Serotypes were assigned with overall high confidence levels by the software (see Dataset S1): 13 were a perfect match to the sequence of reference, 144 had very high confidence, 33 high, 35 good, 11 low, and 19 none. From the strains used in the experiments, two were a perfect match to reference strains, 20 had very high confidence, 3 were high, 7 good and 3 for which the assignment had low confidence.

Bacteriocin detection

We checked for the presence of bacteriocins and other bacterial toxins using BAGEL4 [97]. The results are reported in Table S1.

Depolymerase detection

We checked for the presence of 14 different HMM profiles associated with bacteriophage-encoded depolymerases from multiple bacterial species [98], Table S2. The profiles were matched against the complete set of prophage proteins using HMMER v3.1b2, filtering by the *e*-value of the best domain (maximum 10^{-3}) and the coverage of the profile (minimum 30%). Five additional sequences of depolymerases, validated experimentally in lytic phages of *Klebsiella* (see references in Table S1), were also matched against our dataset using *blastp* (version 2.7.1+, default parameters, and filtering by the *e*-value (maximum 10^{-5}), identity (40%), and coverage (40%).

Identification of CRISPR arrays and R-M systems

(i) *CRISPR-Cas arrays*. We used CRISPRCasFinder [99] (v4.2.18, default parameters) to identify the CRISPR arrays

in all *Klebsiella* genomes used in this study. We excluded arrays with less than three spacers. We then matched each spacer sequence in each array with the complete prophage nucleotide sequences *(blastn* version 2.7.1+, with the -task blastn-short option). Only matches with a spacer coverage of at least 90%, a maximum *e*-value of 10^{-5} and a minimum nucleotide identity of 90% were retained. The resulting matches indicate prophages that are targeted by these spacers, and the full set of these results are presented in Dataset S1. (ii) *R-M systems* were identified using the highly specific and publicly available HMM profiles in https://gitlab.pasteur.fr/erocha/RMS_scripts. If a single protein matched multiple systems, the best hit for each protein-R-M pair was selected. To assess the likelihood that a strain can defend itself against infection by prophages induced from another strain using R-M systems, we calculated the similarity between these proteins (all versus all) using BLAST *(blastp* version 2.7.1+, filtering by identity >50%, *e*-value < 0.0001). We then considered that two R-M associated proteins can target similar recognition sites if their identity is either at above 50% (for Type II and IV REases), 55% (for Type IIG), 60% (for Type II MTases)), or 80% (for Type I and III MTases and Type I and III REases), according to [100]. We inferred the recognition sites targeted by these systems using REBASE (http://rebase.neb.com/rebase/rebase.html, default parameters). We selected the recognition site associated to the protein with the best score, and further chose those whose identity obeyed the thresholds enumerated above (for each type of RMS). The nucleotide motifs associated with these recognition sites were searched for in the intact prophage genomes from the 35 bacterial strains using the Fuzznuc program from the EMBOSS suite (version 6.6.0), with the default parameters, using the option to search the complementary strand when the motif was not a palindrome. Finally, for each pair of bacterial strains A–B (where strain A produces the lysate and strain B is used as a bacterial lawn), we looked for prophages in strain A whose genomes contain recognition sites targeted by R-M systems present in strain B and absent from strain A. If a recognition site was found in any prophage from strain A, we consider that prophage to be putatively targeted by (at least) one R-M defense system of strain B. Because some R-M systems require two (similar) recognition sites in order to effectively target incoming DNA [101], we also did a separate analysis where we require that each motif is found twice in each individual prophage genome. This analysis resulted in similar qualitative results.

Prophage experiments

(i) *Growth curves:* 200 μL of diluted overnight cultures of *Klebsiella* spp. (1:100 in fresh LB) were distributed in a 96-well plate. Cultures were allowed to reach $OD_{600} = 0.2$ and either MMC to 0, 1, or 3 μg/mL or PEG-precipitated induced and filtered supernatants at different PFU/ml was added. Growth was then monitored until late stationary phase. (ii) *PEG-precipitation of virions*. Overnight cultures were diluted 1:500 in fresh LB and allowed to grow until $OD_{600} = 0.2$. MMC was added to final 5 μg/mL. In parallel, non-induced cultures were grown. After 4 h at 37 °C, cultures were centrifuged at 4000 rpm and the supernatant was filtered through 0.22 um. Filtered supernatants were mixed with chilled PEG-NaCl 5X (PEG 8000 20% and 2.5 M of NaCl) and mixed through inversion. Virions were allowed to precipitate for 15 min and pelleted by centrifugation 10 min at 13,000 rpm at 4 °C. The pellets were dissolved in TBS (Tris Buffer Saline, 50 mM Tris-HCl, pH 7.5, 150 mM NaCl). (iii) *All-against-all infection*. Overnight cultures of all strains were diluted 1:100 and allowed to grow until $OD_{600} = 0.8$. 1 mL of bacterial cultures were mixed with 12 mL of top agar (0.7% agar), and 3 mL of the mixture was poured onto a prewarmed LB plate and allowed to dry. Ten microliters of freshly prepared and PEG-precipitated lysates were spotted on the top agar and allowed to grow for 4 h at 37° prior to assessing clearance of bacterial cultures. This was repeated in three independent temporal blocks. (iv) *Calculating plaque forming units (PFU)*. Overnight cultures of sensitive strains were diluted 1:100 and allowed to grow until $OD_{600} = 0.8$. 250 μL of bacterial cultures were mixed with 3 mL of top agar (0.7% agar) and poured intro prewarmed LB plates. Plates were allowed to dry before spotting serial dilutions of induced and non-induced PEG-precipitate virions. Plates were left overnight at room temperature and phage plaques were counted.

Evolution experiment

Three independent clones of each strain (#54, #26, and #36) were used to initiate each evolving of the three evolving populations in four different environments: (i) LB, (ii) LB supplemented with 0.2% citrate, (iii) LB with mytomycin C (MMC, 0.1 μg/mL), and (iv) LB with MMC (0.1 μg/mL) and supplemented with 0.2% citrate. Populations were allowed to grow for 24 h at 37°. Each day, populations were diluted 1:100 and plated on LB to count for capsulated and non-capsulated phenotypes. This experiment was performed in two different temporal blocks and its results combined.

wGRR calculations and network building

We searched for sequence similarity between all proteins of all phages using mmseqs2 [96] with the sensitivity parameter set at 7.5. The results were converted to the blast format for analysis and were filtered with the following parameters: *e*-value lower than 0.0001, at least 35% identity

between amino acids, and a coverage of at least 50% of the proteins. The filtered hits were used to compute the set of bi-directional best hits between each phage pair. This was then used to compute a score of gene repertoire relatedness for each pair of phage genomes, weighted by sequence identity, computed as following:

$$wGRR = \frac{\sum_i^p \mathrm{id}(A_i, B_i)}{\min(\#A, \#B)},$$

where A_i and B_i is the pair i of homologous proteins present in A and B (containing respectively $\#A$ and $\#B$ proteins and p homologs), $\mathrm{id}(A_i,B_i)$ is the percent sequence identity of their alignment, and $\min(\#A,\#B)$ is the total number of proteins of the smallest prophage, the one encoding the smallest number of proteins (A or B). wGRR varies between zero and one. It amounts to zero if there are no orthologs between the elements, and one if all genes of the smaller phage have an ortholog 100% identical in the other phage. Hence, the wGRR accounts for both frequency of homology and degree of similarity among homologs. For instance, three homologous genes with 100% identity between two phages, where the one with the smallest genome is 100 proteins long, would result in a wGRR of 0.03. The same wGRR value would be obtained with six homologous genes with 50% identity. The phage network was built with these wGRR values, using the *networkx* and *graphviz* Phyton (v2.7) packages, and the *neato* algorithm.

Generation of capsule mutant

Isogenic capsule mutants were generated by an in-frame knock-out deletion of gene *wza* (strain #26, #24, #56, #37) or gene *wcaJ* (strain #57 and #58) by allelic exchange. Upstream and downstream sequences of the *wza* or *wcaJ* gene (>500 pb) were amplified using Phusion Master Mix then joined by overlap PCR. All primers used are listed in Table S5. The PCR product was purified using the QIAquick Gel Extraction Kit after electrophoresis in agarose gel 1% and then cloned with the Zero Blunt® TOPO® PCR Cloning Kit (Invitrogen) into competent *E. coli* DH5α strain. KmR colonies were isolated and checked by PCR. A positive Zero Blunt® TOPO® plasmid was extracted using the QIAprep Spin Miniprep Kit, digested for 2 h at 37 °C with ApaI and SpeI restriction enzymes and ligated with T4 DNA ligase overnight to digested pKNG101 plasmid. The ligation was transformed into competent *E. coli* DH5α pir strain, and again into *E. coli* MFD λ−pir strain [102], which was used as a donor strain for conjugation into *Klebsiella* spp. Conjugations were performed for 24 h at 37°. Single cross-over mutants (transconjugants) were selected on Streptomycin plates (200 μg/mL). Plasmid excision was performed after a 48 h culture incubated at 25 °C and double cross-over mutants were selected on LB without salt plus 5% sucrose

at the room temperature. To confirm the loss of the plasmid, colonies were tested for their sensitivity to streptomycin and mutants were confirmed by PCR across the deleted region and further verified by Illumina sequencing.

Data availability

Raw data are available in Dryad: https://doi.org/10.5061/dryad.qfttdz0f3.

Compliance with ethical standards

References

1. Canchaya C, Proux C, Fournous G, Bruttin A, Brussow H. Prophage genomics. Microbiol Mol Biol Rev. 2003;67:238–76.
2. Lu MJ, Henning U. Superinfection exclusion by T-even-type coliphages. Trends Microbiol. 1994;2:137–9.
3. Susskind MA, Wright A, Botstein D. Superinfection exclusion by P22 prophage in lysogens of *Salmonella typhimurium*. IV. genetics and physiology of sieB exclusion. Virology. 1974;62:367–84.
4. Asadulghani M, Ogura Y, Ooka T, Itoh T, Sawaguchi A, Iguchi A, et al. The defective prophage pool of *Escherichia coli* O157: prophage-prophage interactions potentiate horizontal transfer of virulence determinants. PLoS Pathog. 2009;5:e1000408.
5. Matos RC, Lapaque N, Rigottier-Gois L, Debarbieux L, Meylheuc T, Gonzalez-Zorn B, et al. *Enterococcus faecalis* prophage

dynamics and contributions to pathogenic traits. PLoS Genet. 2013;9:e1003539.

6. Touchon M, Bobay LM, Rocha EP. The chromosomal accommodation and domestication of mobile genetic elements. Curr Opin Microbiol. 2014a;22:22–9.

7. Nakayama K, Takashima K, Ishihara H, Shinomiya T, Kageyama M, Kanaya S, et al. The R-type pyocin of *Pseudomonas aeruginosa* is related to P2 phage, and the F-type is related to lambda phage. Mol Microbiol. 2000;38:213–31.

8. Winstanley C, Langille MG, Fothergill JL, Kukavica-Ibrulj I, Paradis-Bleau C, Sanschagrin F, et al. Newly introduced genomic prophage islands are critical determinants of *in vivo* competitiveness in the Liverpool Epidemic Strain of *Pseudomonas aeruginosa*. Genome Res. 2009;19:12–23.

9. Bossi L, Fuentes JA, Mora G, Figueroa-Bossi N. Prophage contribution to bacterial population dynamics. J Bacteriol. 2003; 185:6467–71.

10. Fortier LC, Sekulovic O. Importance of prophages to evolution and virulence of bacterial pathogens. Virulence. 2013;4:354–65.

11. Nanda AM, Thormann K, Frunzke J. Impact of spontaneous prophage induction on the fitness of bacterial populations and host-microbe interactions. J Bacteriol. 2015;197:410–9.

12. Brown SP, Le Chat L, De Paepe M, Taddei F. Ecology of microbial invasions: amplification allows virus carriers to invade more rapidly when rare. Curr Biol. 2006;16:2048–52.

13. Sousa JAM, Rocha EPC. Environmental structure drives resistance to phages and antibiotics during phage therapy and to invading lysogens during colonisation. Sci Rep. 2019;9:3149.

14. Roux S, Hallam SJ, Woyke T, Sullivan MB. Viral dark matter and virus-host interactions resolved from publicly available microbial genomes. Elife. 2015;4:e08490.

15. Touchon M, Bernheim A, Rocha EP. Genetic and life-history traits associated with the distribution of prophages in bacteria. ISME J. 2016;10:2744–54.

16. Bobay LM, Rocha EP, Touchon M. The adaptation of temperate bacteriophages to their host genomes. Mol Biol Evol. 2013; 30:737–51.

17. Wagner PL, Waldor MK. Bacteriophage control of bacterial virulence. Infect Immun. 2002;70:3985–93.

18. Chen J, Quiles-Puchalt N, Chiang YN, Bacigalupe R, Fillol-Salom A, Chee MSJ, et al. Genome hypermobility by lateral transduction. Science. 2018;362:207–12.

19. Touchon M, Moura de Sousa JA, Rocha EP. Embracing the enemy: the diversification of microbial gene repertoires by phage-mediated horizontal gene transfer. Curr Opin Microbiol. 2017;38:66–73.

20. Haaber J, Leisner JJ, Cohn MT, Catalan-Moreno A, Nielsen JB, Westh H, et al. Bacterial viruses enable their host to acquire antibiotic resistance genes from neighbouring cells. Nat Commun. 2016;7:13333.

21. Brisse S, Grimont F, Grimont PAD. The genus *Klebsiella*. The Prokaryotes. New York, USA: Springer; 2006. p. 159–96.

22. Lee CR, Lee JH, Park KS, Jeon JH, Kim YB, Cha CJ, et al. Antimicrobial resistance of hypervirulent *Klebsiella pneumoniae*: epidemiology, hypervirulence-associated determinants, and resistance mechanisms. Front Cell Infect Microbiol. 2017;7:483.

23. Navon-Venezia S, Kondratyeva K, Carattoli A. *Klebsiella pneumoniae*: a major worldwide source and shuttle for antibiotic resistance. FEMS Microbiol Rev. 2017;41:252–75.

24. Blin C, Passet V, Touchon M, Rocha EPC, Brisse S. Metabolic diversity of the emerging pathogenic lineages of *Klebsiella pneumoniae*. Environ Microbiol. 2017;19:1881–98.

25. Wyres KL, Wick RR, Judd LM, Froumine R, Tokolyi A, Gorrie CL, et al. Distinct evolutionary dynamics of horizontal gene transfer in drug resistant and virulent clones of *Klebsiella pneumoniae*. PLoS Genet. 2019;15:e1008114.

26. Holt KE, Wertheim H, Zadoks RN, Baker S, Whitehouse CA, Dance D, et al. Genomic analysis of diversity, population structure, virulence, and antimicrobial resistance in *Klebsiella pneumoniae*, an urgent threat to public health. Proc Natl Acad Sci USA. 2015;112:E3574–81.

27. Mori M, Ohta M, Agata N, Kido N, Arakawa Y, Ito H, et al. Identification of species and capsular types of *Klebsiella* clinical isolates, with special reference to *Klebsiella planticola*. Microbiol Immunol. 1989;33:887–95.

28. Wyres KL, Wick RR, Gorrie C, Jenney A, Follador R, Thomson NR, et al. Identification of *Klebsiella* capsule synthesis loci from whole genome data. Micro Genom. 2016;2:e000102.

29. Pan YJ, Lin TL, Chen CT, Chen YY, Hsieh PF, Hsu CR, et al. Genetic analysis of capsular polysaccharide synthesis gene clusters in 79 capsular types of *Klebsiella* spp. Sci Rep. 2015;5:15573.

30. Favre-Bonte S, Licht TR, Forestier C, Krogfelt KA. *Klebsiella pneumoniae* capsule expression is necessary for colonization of large intestines of streptomycin-treated mice. Infect Immun. 1999;67:6152–6.

31. Alvarez D, Merino S, Tomas JM, Benedi VJ, Alberti S. Capsular polysaccharide is a major complement resistance factor in lipo-polysaccharide O side chain-deficient *Klebsiella pneumoniae* clinical isolates. Infect Immun. 2000;68:953–5.

32. Campos MA, Vargas MA, Regueiro V, Llompart CM, Alberti S, Bengoechea JA. Capsule polysaccharide mediates bacterial resistance to antimicrobial peptides. Infect Immun. 2004;72:7107–14.

33. Doorduijn DJ, Rooijakkers SHM, van Schaik W, Bardoel BW. Complement resistance mechanisms of *Klebsiella pneumoniae*. Immunobiology. 2016;221:1102–9.

34. Rendueles O, Garcia-Garcera M, Neron B, Touchon M, Rocha EPC. Abundance and co-occurrence of extracellular capsules increase environmental breadth: Implications for the emergence of pathogens. PLoS Pathog. 2017;13:e1006525.

35. Rendueles O, de Sousa JAM, Bernheim A, Touchon M, Rocha EPC. Genetic exchanges are more frequent in bacteria encoding capsules. PLoS Genet. 2018;14:e1007862.

36. Chewapreecha C, Harris SR, Croucher NJ, Turner C, Marttinen P, Cheng L, et al. Dense genomic sampling identifies highways of pneumococcal recombination. Nat Genet. 2014;46:305–9.

37. Moller AG, Lindsay JA, Read TD. Determinants of phage host range in *Staphylococcus* species. Appl Environ Microbiol. 2019;85:e00209–19.

38. Negus D, Burton J, Sweed A, Gryko R, Taylor PW. Poly-gamma-(D)-glutamic acid capsule interferes with lytic infection of *Bacillus anthracis* by *B. anthracis*-specific bacteriophages. Appl Environ Microbiol. 2013;79:714–7.

39. Scholl D, Adhya S, Merril C. *Escherichia coli* K1's capsule is a barrier to bacteriophage T7. Appl Environ Microbiol. 2005;71:4872–4.

40. Niemann H, Frank N, Stirm S. *Klebsiella* serotype-13 capsular polysaccharide: primary structure and depolymerization by a bacteriophage-borne glycanase. Carbohydr Res. 1977a;59:165–77.

41. Niemann H, Kwiatkowski B, Westphal U, Stirm S. *Klebsiella* serotype 25 capsular polysaccharide: primary structure and depolymerization by a bacteriophage-borne glycanase. J Bacteriol. 1977b;130:366–74.

42. Pan YJ, Lin TL, Chen CC, Tsai YT, Cheng YH, Chen YY, et al. *Klebsiella* phage PhiK64-1 encodes multiple depolymerases for multiple host capsular types. J Virol. 2017;91:e02457–02416.

43. Bessler W, Freund-Molbert E, Knufermann H, Rduolph C, Thurow H, Stirm S. A bacteriophage-induced depolymerase active on *Klebsiella* K11 capsular polysaccharide. Virology. 1973;56:134–51.

44. Thurow H, Niemann H, Rudolph C, Stirm S. Host capsule depolymerase activity of bacteriophage particles active on *Klebsiella* K20 and K24 strains. Virology. 1974;58:306–9.

45. Latka A, Maciejewska B, Majkowska-Skrobek G, Briers Y, Drulis-Kawa Z. Bacteriophage-encoded virion-associated enzymes to overcome the carbohydrate barriers during the infection process. Appl Microbiol Biotechnol. 2017;101:3103–19.

46. Scholl D, Rogers S, Adhya S, Merril CR. Bacteriophage K1-5 encodes two different tail fiber proteins, allowing it to infect and replicate on both K1 and K5 strains of *Escherichia coli*. J Virol. 2001;75:2509–15.

47. Lin TL, Hsieh PF, Huang YT, Lee WC, Tsai YT, Su PA, et al. Isolation of a bacteriophage and its depolymerase specific for K1 capsule of *Klebsiella pneumoniae*: implication in typing and treatment. J Infect Dis. 2014;210:1734–44.

48. Pan YJ, Lin TL, Chen YY, Lai PH, Tsai YT, Hsu CR, et al. Identification of three podoviruses infecting *Klebsiella* encoding capsule depolymerases that digest specific capsular types. Micro Biotechnol. 2019;12:472–86.

49. Bertozzi Silva J, Storms Z, Sauvageau D. Host receptors for bacteriophage adsorption. FEMS Microbiol Lett. 2016;363:fnw002.

50. Arndt D, Grant JR, Marcu A, Sajed T, Pon A, Liang Y, et al. PHASTER: a better, faster version of the PHAST phage search tool. Nucleic Acids Res. 2016;44:W16–21.

51. Daubin V, Lerat E, Perriere G. The source of laterally transferred genes in bacterial genomes. Genome Biol. 2003;4:R57.

52. Rocha EP, Danchin A. Base composition bias might result from competition for metabolic resources. Trends Genet. 2002;18:291–4.

53. Bobay LM, Touchon M, Rocha EP. Pervasive domestication of defective prophages by bacteria. Proc Natl Acad Sci USA. 2014;111:12127–32.

54. Casjens SR, Gilcrease EB, Huang WM, Bunny KL, Pedulla ML, Ford ME, et al. The pKO2 linear plasmid prophage of *Klebsiella oxytoca*. J Bacteriol. 2004;186:1818–32.

55. Leclercq S, Cordaux R. Do phages efficiently shuttle transposable elements among prokaryotes? Evolution. 2011;65:3327–31.

56. Ghazaryan L, Tonoyan L, Ashhab AA, Soares MI, Gillor O. The role of stress in colicin regulation. Arch Microbiol. 2014;196:753–64.

57. Labrie SJ, Samson JE, Moineau S. Bacteriophage resistance mechanisms. Nat Rev Microbiol. 2010;8:317–27.

58. Samson JE, Magadan AH, Sabri M, Moineau S. Revenge of the phages: defeating bacterial defences. Nat Rev Microbiol. 2013;11:675–87.

59. Wyres KL, Gorrie C, Edwards DJ, Wertheim HF, Hsu LY, Van Kinh N, et al. Extensive capsule locus variation and large-scale genomic recombination within the *Klebsiella pneumoniae* clonal group 258. Genome Biol Evol. 2015;7:1267–79.

60. Tan D, Zhang Y, Qin J, Le S, Gu J, Chen LK et al. A frameshift mutation in *wcaJ* associated with phage resistance in *Klebsiella pneumoniae*. Microorganisms. 2020;8:378.

61. Buffet A, Rocha EPC, Rendueles O. Selection for the bacterial capsule in the absence of biotic and abiotic aggressions depends on growth conditions. BioRxiv. 2020. https://doi.org/10.1101/2020.04.27.059774.

62. Julianelle LA. Bacterial variation in cultures of Friedlander's *Bacillus*. J Exp Med. 1928;47:889–902.

63. Randall WA. Colony and antigenic variation in *Klebsiella pneumoniae* types A, B and C. J Bacteriol. 1939;38:461–77.

64. Flores CO, Meyer JR, Valverde S, Farr L, Weitz JS. Statistical structure of host-phage interactions. Proc Natl Acad Sci USA. 2011;108:E288–97.

65. Mathieu A, Dion M, Deng L, Tremblay D, Moncaut E, Shah SA, et al. Virulent coliphages in 1-year-old children fecal samples are fewer, but more infectious than temperate coliphages. Nat Commun. 2020;11:385.

66. Weitz JS, Poisot T, Meyer JR, Flores CO, Valverde S, Sullivan MB, et al. Phage-bacteria infection networks. Trends Microbiol. 2013;21:82–91.

67. Brueggemann AB, Harrold CL, Rezaei Javan R, van Tonder AJ, McDonnell AJ, Edwards BA. Pneumococcal prophages are diverse, but not without structure or history. Sci Rep. 2017;7:42976.

68. Castillo D, Kauffman K, Hussain F, Kalatzis P, Rorbo N, Polz MF, et al. Widespread distribution of prophage-encoded virulence factors in marine *Vibrio* communities. Sci Rep. 2018;8:9973.

69. Allen HK, Looft T, Bayles DO, Humphrey S, Levine UY, Alt D, et al. Antibiotics in feed induce prophages in swine fecal microbiomes. mBio. 2011;2:e00260–11.

70. Otsuji N, Sekiguchi M, Iijima T, Takagi Y. Induction of phage formation in the lysogenic *Escherichia coli* K-12 by mitomycin C. Nature. 1959;184:1079–80.

71. Comeau AM, Tetart F, Trojet SN, Prere MF, Krisch HM. Phage-antibiotic synergy (PAS): beta-lactam and quinolone antibiotics stimulate virulent phage growth. PLoS One. 2007;2:e799.

72. Kim M, Jo Y, Hwang YJ, Hong HW, Hong SS, Park K et al. Phage-antibiotic synergy via delayed lysis. Appl Environ Microbiol. 2018;84:e02085–18.

73. De Paepe M, Tournier L, Moncaut E, Son O, Langella P, Petit MA. Carriage of lambda latent virus is costly for its bacterial host due to frequent reactivation in monoxenic mouse intestine. PLoS Genet. 2016;12:e1005861.

74. Phanphak S, Georgiades P, Li R, King J, Roberts IS, Waigh TA. Super-resolution fluorescence microscopy study of the production of K1 capsules by *Escherichia coli*: evidence for the differential distribution of the capsule at the poles and the equator of the cell. Langmuir. 2019;35:5635–46.

75. Krinos CM, Coyne MJ, Weinacht KG, Tzianabos AO, Kasper DL, Comstock LE. Extensive surface diversity of a commensal microorganism by multiple DNA inversions. Nature. 2001;414:555–8.

76. Tzeng YL, Thomas J, Stephens DS. Regulation of capsule in *Neisseria meningitidis*. Crit Rev Microbiol. 2016;42:759–72.

77. Bondy-Denomy J, Qian J, Westra ER, Buckling A, Guttman DS, Davidson AR, et al. Prophages mediate defense against phage infection through diverse mechanisms. ISME J. 2016;10:2854–66.

78. Zhang YF, LeJeune JT. Transduction of bla(CMY-2), tet(A), and tet(B) from *Salmonella enterica* subspecies enterica serovar Heidelberg to S-Typhimurium. Vet Microbiol. 2008;129:418–25.

79. Brussow H, Kutter E. Phage ecology. Bacteriophages: biology and application. Boca Raton, Florida: CRC Press; 2005. p. 129–64.

80. Calef E, Marchelli C, Guerrini F. The formation of superinfection-double lysogens of phage lambda in *Escherichia coli* K-12. Virology. 1965;27:1–10.

81. Scott JR, West BW, Laping JL. Superinfection immunity and prophage repression in phage P1. IV. The c1 repressor bypass function and the role of c4 repressor in immunity. Virology. 1978;85:587–600.

82. Bakk A, Metzler R. Nonspecific binding of the OR repressors CI and Cro of bacteriophage lambda. J Theor Biol. 2004;231:525–33.

83. Casjens S. Prophages and bacterial genomics: what have we learned so far? Mol Microbiol. 2003;49:277–300.

84. Hendrix RW, Smith MCM, Burns RN, Ford ME, Hatfull GF. Evolutionary relationships among diverse bacteriophages and prophages: All the world's a phage. Proc Natl Acad Sci USA. 1999;96:2192–7.

85. Ramisetty BCM, Sudhakari PA. Bacterial 'Grounded' Prophages: hotspots for genetic renovation and innovation. Front Genet. 2019;10:65.

86. Venturini C, Ben Zakour N, Bowring B, Morales S, Cole R, Kovach Z et al. *K. pneumoniae* ST258 genomic variability and bacteriophage susceptibility. bioRxiv. 2019. https://doi.org/10.1101/628339.

87. Hsieh PF, Lin HH, Lin TL, Chen YY, Wang JT. Two T7-like bacteriophages, K5-2 and K5-4, each encodes two capsule depolymerases: isolation and functional characterization. Sci Rep. 2017;7:4624.

88. Majkowska-Skrobek G, Latka A, Berisio R, Squeglia F, Maciejewska B, Briers Y, et al. Phage-borne depolymerases decrease *Klebsiella pneumoniae* resistance to innate defense mechanisms. Front Microbiol. 2018;9:00209–19.

89. Paczosa MK, Mecsas J. *Klebsiella pneumoniae*: going on the offense with a strong defense. Microbiol Mol Biol Rev. 2016;80: 629–61.

90. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics. 2014;30:2068–9.

91. Grazziotin AL, Koonin EV, Kristensen DM. Prokaryotic virus orthologous groups (pVOGs): a resource for comparative genomics and protein family annotation. Nucleic Acids Res. 2017;45: D491–D498.

92. Eddy SR. Accelerated profile HMM searches. PLoS Comput Biol. 2011;7:e1002195.

93. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. BMC Bioinforma. 2009;10:421.

94. Siguier P, Perochon J, Lestrade L, Mahillon J, Chandler M. ISfinder: the reference centre for bacterial insertion sequences. Nucleic Acids Res. 2006;34:D32–6.

95. Touchon M, Cury J, Yoon EJ, Krizova L, Cerqueira GC, Murphy C, et al. The genomic diversification of the whole Acinetobacter genus: origins, mechanisms, and consequences. Genome Biol Evol. 2014b;6:2866–82.

96. Steinegger M, Soding J. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. Nat Biotechnol. 2017;35:1026–8.

97. van Heel AJ, de Jong A, Song C, Viel JH, Kok J, Kuipers OP. BAGEL4: a user-friendly web server to thoroughly mine RiPPs and bacteriocins. Nucleic Acids Res. 2018; 46:W278–81.

98. Pires DP, Oliveira H, Melo LD, Sillankorva S, Azeredo J. Bacteriophage-encoded depolymerases: their diversity and biotechnological applications. Appl Microbiol Biotechnol. 2016; 100:2141–51.

99. Couvin D, Bernheim A, Toffano-Nioche C, Touchon M, Michalik J, Neron B, et al. CRISPRCasFinder, an update of CRISRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins. Nucleic Acids Res. 2018;46:W246–51.

100. Oliveira PH, Touchon M, Rocha EP. Regulation of genetic flux between bacteria by restriction-modification systems. Proc Natl Acad Sci USA. 2016;113:5658–63.

101. Mucke M, Kruger DH, Reuter M. Diversity of Type II restriction endonucleases that require two DNA recognition sites. Nucleic Acids Res. 2003;31:6079–84.

102. Ferrieres L, Hemery G, Nham T, Guerout AM, Mazel D, Beloin C, et al. Silent mischief: bacteriophage Mu insertions contaminate products of *Escherichia coli* random mutagenesis performed using suicidal transposon delivery plasmids mobilized by broad-host-range RP4 conjugative machinery. J Bacteriol. 2010;192:6418–27.

103. Guy L, Kultima JR, Andersson SG. genoPlotR: comparative gene and genome visualization in R. Bioinformatics. 2010;26: 2334–5.