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

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

[1] Microbial Evolutionary Genomics, Institut Pasteur, CNRS, UMR3525, Paris, 75015, France
[2] Ecole Doctoral FIRE – Programme Bettencourt, CRIS, Paris, France
* olaya.rendueles-garcia@pasteur.fr

Running title: Prophages are serotype-specific in *Klebsiella*

Keywords: Capsule / Serotype / *Klebsiella* / Prophages / Lysogeny

ABSTRACT

Klebsiella species are able to colonize a wide range of environments and include daunting 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 *Klebsiella* 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, most interactions are null, and is modular because of the capsule serotype. Capsule mutants remain uninfected showing that the capsule is required for successful infection. Accordingly, experiments where bacteria are predated by their own prophages result in accelerated loss of the capsule. Our results show that 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. This should also restrict phage-driven gene flow across the species. However, if capsule loss driven by phage predation is followed by the acquisition of a novel capsule, then bacteria shift to another module where they can be infected by other phages that can transduce genes from other bacterial strains.

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. On the contrary, temperate phages, which can either follow a lytic cycle, or integrate into the host genome and produce a lysogen, have been comparatively less studied. These integrated phages, hereafter referred to as prophages, replicate vertically with the host and protect them from infections by similar phages (Canchaya et al 2003, Lu and Henning 1994, Susskind et al 1974). Most prophage genes are silent and have little impact in bacterial fitness as long as there is no induction of the lytic cycle (Canchaya et al 2003). 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 (Asadulghani et al 2009, Ventura et al 2004). Upon induction, some of them cannot excise (cryptic prophages), replicate or infect. In contrast, intact prophages can be induced (by either extrinsic or intrinsic factors) and resume a lytic cycle, producing viable viral particles. Prophage inactivation and further domestication may lead to the co-option of some phage functions by the bacterial host (Touchon et al 2014a). For instance, some bacteriocins result from the domestication of phage tails (Nakayama et al 2000, Winstanley et al 2009).

Around half of the bacteria are lysogens and some genomes can have up to 20 different prophages (Roux et al 2015, Touchon et al 2016). The frequency of prophages is higher in bacteria with larger genomes, in pathogens, and in fast-growing bacteria (Touchon et al 2016). In *E. coli,* a third of its pangenome, or the diversity of its genetic repertoire, lies within phages or their remnants (Bobay et al 2013). This has important consequences, as a number of studies have highlighted how prophages can affect bacterial population dynamics and evolution (Bossi et al 2003, Fortier and Sekulovic 2013, Nanda et al 2015). This is due to two key traits: their ability to kill bacteria lacking the prophage and to drive horizontal gene transfer between bacteria. Induction of prophages by a small subset of a population produces virions that can infect susceptible bacteria and thus facilitate colonization (Bossi et al 2003, Brown et al 2006, Sousa and Rocha 2019). Prophages are part of the bacterial genome and the

expression of their accessory genes may result in phenotypic changes. This is the case in some important pathogens that require virulence factors encoded in prophages (Wagner and Waldor 2002). Prophages may also facilitate horizontal transfer between bacteria by one of several mechanisms of transduction (Canchaya et al 2003, Chen et al 2018, Touchon et al 2017). Interestingly, bacterial populations can acquire adaptive genes from their competitors by killing them with induced prophages and recovering their genes by transduction (Haaber et al 2016). While several experimental studies have detailed these mechanisms, it is yet unclear how temperate phages shape the diversity of bacterial lysogens.

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, and mammals (Brisse et al 2006). Further, *Klebsiella spp.* can cause various diseases such as urinary tract infections, acute liver abscesses, pneumonia, infectious wounds and dental infections (Lee et al 2017, Navon-Venezia et al 2017). They commonly cause severe hospital outbreaks associated with multidrug resistance (MDR), 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 (Blin et al 2017), partly acquired by horizontal gene transfer (Holt et al 2015, Navon-Venezia et al 2017, Wyres et al 2019). Additionally, *Klebsiella spp*. code for an extracellular capsule that is highly variable within 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 *Klebsiella pneumoniae* there are 77 serologically-defined serotypes (Mori et al 1989) and more than 130 computationally (Pan et al 2015, Wyres et al 2016). The capsule is considered a major virulence factor, it is required, amongst others, for intestinal colonization (Favre-Bonte et al 1999), and has been associated with resistance to the immune response and to antibiotics (Alvarez et al 2000, Campos et al 2004, Doorduijn et al 2016). From an ecological point of view, the capsule is associated with bacteria able to colonize diverse environments (Rendueles et al 2017, Rendueles et al 2018). 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 (Rendueles et al 2018). This is surprising because capsules were proposed to decrease gene flow (Chewapreecha et al 2014) and some phages are known to be blocked by the capsule (Moller et al 2019, Negus et al 2013, Scholl et al 2005). However, several lytic phages of *Klebsiella* are known to have depolymerase activity in their tail fibers (Bessler et al 1973, Niemann et al 1977a, Niemann et al 1977b, Pan et al 2017, Thurow et al 1974). These depolymerases specifically digest oligosaccharidic bonds in certain capsules (Latka et al 2017) and allow phages to access the outer membrane (Scholl et al 2001). Since depolymerases are specific to one or a few capsule types (Lin et al 2014, Thurow et al 1974), this implicates that some phages interact with capsules in a serotype-specific manner (Pan et al 2017, Pan et al 2019, Thurow et al 1974). To date, the role of prophages in the intra-specific interactions and their impact in *Klebsiella*'s population biology is not well understood because the distribution of prophages in the genomes of *Klebsiella* has not been systematically studied.

Klebsiella are interesting models to study the role of prophages, because of the interplay between colonization, the capsule and phage infections. 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 (Arndt et al 2016) 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%, Figure 1AB, Figure S1). The complete list of genomes and prophages is available in Dataset S1. Most of the genomes were poly-lysogenic, encoding more than one prophage per genome (Figure S1). However, the number of prophages varied markedly across genomes, ranging from one to 16, with a median of 6 per genome. Additionally, the total number of prophages in *Klebsiella* spp. varied significantly across species (Kruskal-Wallis, $df = 6$, $P < 0.001$, Figure 1B). More specifically, the average number of "intact" prophages varied between eight for *K. oxytoca* and less than one for *K. quasipneumoniae subsp. quasipneumoniae*. *Klebsiella pneumoniae*, the most represented species of our dataset (~77% of the genomes), has an average of nine prophages per genome, of which four are classified as "intact" (Figure 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) (Figure S2AB). When compared with the one hundred most sequenced bacterial species, the number of prophages in *Klebsiella* is very high. It ranks *K. pneumoniae* within the 5th percentile of the most prophage-rich species, comparable to *E. coli* and *Yersinia enterocolitica* (Figure S3). This shows that prophages are important contributors for 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, Figure S2B), as typically observed for horizontally transferred genetic elements (Daubin et al 2003, Rocha and Danchin 2002). Their length varies from 13 to 137 kb, for an average of 46kb. Since temperate dsDNA phages of enterobacteria are usually larger than 30kb (Bobay et al 2014), this suggests that a small

fraction of the prophages might be incomplete (Figure S2C). Among the "intact" prophages detected in the 35 strains analysed from our laboratory collection (Dataset S1 and Figure S1) and isolated from different environments and representative of the *Klebsiella* genetic diversity (Blin et al 2017), we chose nine to characterize in detail in terms of genetic architecture (Figure 2). A manual and computational search for recombination sites (*att*) in these nine prophages (See Methods, Figure 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 nine prophages. One of these prophages (#62(2)) had a protein of *ca.* 4200 aa (Figure 2), homologous to the tail component (gp21) of the linear phage-plasmid phiKO2 of *Klebsiella oxytoca* (Casjens et al 2004). The prophages had integrases and were flanked by recombination sites, suggesting that they could still be able to excise themselves from the chromosome (Figure 2). Nevertheless, some prophages encoded a small number of insertion sequences (IS), which are expected to accumulate and degrade prophages(Leclercq and Cordaux 2011).

Klebsiella spp. prophages can be released into the environment

To further characterize the prophages of *Klebsiella,* we sought to experimentally assess their ability to excise and be released into the environment. In the 35 strains analyzed from of our collection (Figure 3), a total of 95 prophages were detected, among which 40 were classed as "intact" (including the nine prophages whose genomic composition is characterized in Figure 2). Eleven strains had no "intact" prophages, and four of these also lacked "questionable" prophages.

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 (Figure 3 and Figure 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.* one hour after exposure to MMC (Figure 3 and Figure 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 18 different phages from eleven different strains by PCR, both in induced and non-induced PEG-precipitated filter-sterilized supernatants (see Methods). Indeed, all 18 phages were amplified, indicating that viral genomes are released into the environment. As expected, amplification bands were weaker in the non-induced supernatant compared to the induced (Figure S5A). We further verified the recircularization of these prophages in induced supernatants. We detected the recircularization of eight phages belonging to six strains (#25, #36, #62, #63, #37 and #50), five of which were classified as "intact", and three as "questionable" by PHASTER (Figure S5B). We did not obtain a clean PCR product for eight phages, despite using different primers and PCR settings. This is most likely due to the aforementioned inaccuracies in the delimitation of the prophages. 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

To characterize the potential of prophages to infect other *Klebsiella spp.*, we produced lysates of all 35 strains by MMC induction and PEG-precipitation of the filter-sterilized supernatants (see Methods). We then tested the ability of all lysates to produce a clearing on bacterial overlays of every strain. This experiment involves 1225 possible lysate-bacteria combinations, with the caveat that some lysates can have more than one type of phage. We observed that 21 strains out of the 35 (60%) produced a lysate that could infect another strain, of which, five (14%) could still be re-infected by their own lysate. However, we only observed clearances, or plaques, in 75 instances (6%) (Figure 4A). Plaques were small and were not surrounded by enlarged halos typically indicative of depolymerase activity. To confirm that these plaques are the result of lysis by phage particles, we observed that growth is arrested in a dose-dependent manner after addition of purified phage to exponentially growing cultures of sensitive strains (Figure S6). Moreover, phage infection was not observed in the presence of citrate, a known inhibitor of phage adsorption (Figure S6). One surprising feature of these experiments is that among the 75 infections, only 17 were observed in the three independent replicates of the experiment (Figure S7). This suggests that despite the

ability of the strains to produce viral particles, re-infections are not very efficient. One possible explanation could be that the resident phages are unfit due to accumulation of deleterious mutations during lysogeny. Alternatively, inefficient infections could also result from stochasticity in capsule production (Krinos et al 2001, Tzeng et al 2016) or uneven distribution of capsule throughout the cell (Phanphak et al 2019). Infections between lysates and target bacteria from the same serotype seem more effective (and reproducible) than the others (Figure S5). Overall, this analysis shows that most lysates cannot infect other strains, producing a very sparse matrix of prophage-bacteria interactions potentially due to the presence of a capsule.

Induced *Klebsiella* prophages can lysogenize other strains

Infection of a bacterium by a temperate phage, induced from other strain, can result again in lysogeny. To test the ability of the 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 capsular serotype as BJ1 (KL2) and we have PCR evidence that they produce viral particles (see above) (Fig. S2B). We grew BJ1 in contact with these two lysates, and from the survivor BJ1 cells, we isolated 93 clones from 3 independently challenged cultures. We re-exposed these clones to a lysate from strain #54, to test if they had become resistant. Almost all clones (96%) of BJ1 could grow normally (Figure 4B), suggesting that they had acquired resistance by lysogenization or some other mechanism. Most of these clones displayed a significant cell death upon exposure to MMC (whereas the ancestral strain was insensitive), indicating that their resistance is mostly driven by lysogenization. Positive PCRs showed that 57 clones, belonging to the three independent BJ1 cultures previously exposed to #54 lysate, had acquired at least one "intact" prophage (Figure 4B). Furthermore, at least four of these 57 clones now carry two different "intact" prophages from #54. We then tested survivor BJ1 cells that had been challenged with the lysate of strain #25 (Figure 4C). In contrast to what we observed in clones challenged with strain #54 lysate, no BJ1 clones showed signs of lysogenization when it was exposed to MMC (and as verified by PCR) (Figure 4C), indicating that most clones became resistant by other mechanisms.

We also exposed *K. quasipneumoniae subsp similipneumoniae* (#38) to a lysate from a *K. quasipneumoniae subsp quasipneumoniae* (#36) that consistently infects #38 even if it belongs to a different subspecies and has a different capsular serotype (Figure 4A). Survivor

clones of strain #38 had been lysogenized showing that phage infection, and lysogenization, could occur across distinct *Klebsiella* subspecies (Figure 4D). We checked by PCR the presence of the four different prophages from strain #36 (Figure S2) in the surviving clones of strain #38. We observed that only one phage of #36 (its genomic structure is shown in Figure 2) lysogenized 61 out 93 surviving clones (Figure 4D). Interestingly, four #36 lysogenized clones did not display cell death when exposed to mytomycin C, suggesting that these phages, which were able to excise in other clones, might have accumulated mutations that permit infection and integration but not reactivation of lytic cycle once integrated in its host. Taken together, our results show that some *Klebsiella* prophages can lysogenize other strains, including from other subspecies.

The capsule plays a major role in shaping phage infections

We aimed to understand the patterns of phage infection across *Klebsiella* strains and the molecular determinants that underlie infection events. We first hypothesized that the presence of other resident phages, similar to those by which they are being challenged, the so-called, superinfection resistance, could shape the infection matrix. Consistently, the most sensitive strain in our panel (#26) does not have any detectable prophages, rendering it sensitive to infection by numerous lysates. If exclusion was the main cause of infection failure, one would expect more frequent infections between more distantly related bacteria, because these are less likely to have prophages similar to the challenging phages. The matrix in Figure 4A reveals the opposite pattern; phages tend to infect very closely related strains of the same species. We have no means of making high-throughput experimental analysis of superinfection resistance. Instead, we tested if resistant strains had very similar prophages. For this, we calculated the network of homology between all detected "intact" prophages in our panel using weighted Gene Repertoire Relatedness (wGRR) (see Methods). The network of the wGRR shows that nine prophages (20%) are singletons (no other prophage with values higher than 50%) and thus unlikely to provide resistance to superinfection (Figure 2 and 5A). Results are qualitatively similar when we include questionable prophages in the analysis (Figure S8). We cannot exclude that prophages within the nine groups (or clusters, Figure 5A) provide cross-resistance, but most combinations are between groups of prophages of different families and cross-resistance does not seem a parsimonious explanation for our observations.

We then tested the hypothesis that the probability of infection of a strain by the lysate of another was determined by their defence systems. We focused on restriction-modification and CRISPR-Cas systems, since these are the most common, the best characterized to date, and those for which detection tools are available (Labrie et al 2010, Samson et al 2013). We searched for restriction-modification systems and Cas systems in the genomes 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 Cas systems in a genome ($P > 0.05$, Wilcox test) or the number of RMS systems (Spearman correlation $P > 0.05$). We thus conclude that there is no evidence for an effect of defence systems in shaping the observed pattern of infections between different strains.

We then tested if the capsule serotype shapes the infection network of the *Klebsiella* prophages. We used Kleborate to serotype the strains from the genome sequence and compared the intra-serotype with the inter-serotype interactions. We observed that thirty-five successful cross-strain infections occurred between lysates and sensitive bacteria from the same serotype, out of a total of 105 possible combinations (33%). In contrast, only 3.6% of the 1120 possible inter-serotype interactions were observed (40 infections in total, Figure 4A, $P<0.0001$, Fisher's exact test). For example, strain #37 was only infected by lysates produced by the other strain of the KL24 serotype (#51), Figure 2). Similarly, strains #57, #55 and #25 were only infected by lysates from strains of the same serotype (KL2) (Figure 4A). Intriguingly, the lysate of one strain alone (#63) produced an inhibition halo in lawns of fifteen strains from different serotype. Genomic analyses of strain #63 revealed the presence of a plasmid-encoded colicin operon (see Methods). It is thus possible that MMC, besides inducing the prophages (Figure S5), also stimulates the release of this bacteriocin, resulting in the killing of many different strains.

The specificity of phages for a given capsule serotype could be explained by the presence of depolymerases in their tail fibers. We searched the genomes of the prophages for published depolymerase domains (see Methods, Table S1), but find very little similarity with any of them (Table S2 and S3). There is also no clear correlation with any specific serotype. Nevertheless, the large tail protein of phage #62 (2) (Figure 2) had duplicated galactose-binding sites. It could potentially have a depolymerase activity against capsules composed of galactose subunits, which fits the observation that its host serotype (KL107) comprises galactose subunits. However, phages from #62 could not infect strain #63, which has the same

capsule serotype suggesting that either this putative depolymerase is not functional or that the phages, although produced (Figure S5), are not infectious. This suggests that the serotype-specificity of infections might be due to mechanisms other than the presence of capsule depolymerases in phages.

It is often mentioned in the literature (Scholl et al 2005) that capsules protect against phages. However, the results above show that induced prophages tend to infect strains of the same serotype, suggesting that strains lacking a capsule would be immune to phages. We tested if isogenic capsule mutants from different serotypes (KL2, KL30 and KL1) were sensitive to lysates. We verified by microscopy and biochemical capsule quantification that these mutants lacked a capsule in the culture conditions and then challenged them with the lysates of all 35 strains. Wild type strains #56 (KL1) and #24 (KL30) are resistant to all lysates, as are their capsule mutants. Capsule mutants from BJ1 (KL2) were resistant to phages, but not the complemented mutants nor the wild-type for phages of the same serotype (Figure 5BCD). Hence, phages targeting BJ1 require the presence of the capsule to infect. These results show that the loss of the capsule does not make bacteria more sensitive to phages. On the contrary, loss of the capsule may result in resistance to phages adapted to that specific serotype. Taken together, our results indicate that phage infections, and thus prophage exchange, require the presence of a capsule in *Klebsiella* and is more efficient between strains with similar capsule serotypes.

Phage predation stimulates capsule loss

If the capsule is necessary for phage infection, then phage predation might lead to the selection of non-capsulated mutants because these would be immune to the phage. To test this hypothesis, we performed a short-term evolution experiment (ten days, ca 70 generations). We assessed the natural emergence of non-capsulated bacteria in 3 different strains: (i) a strain with no temperate phages (BJ1: #26); (ii) a strain that produces phages infecting many strains including itself (#54); and (iii) a strain that is resistant to its own phages (#36). 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 titre, 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, as previously

observed (Julianelle 1928, Randall 1939). This process should be accelerated if phage predation drives selection for the non-capsulated mutants (presence of phages, absence of citrate). 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 (Figure 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 (Figure 6B). The 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 (MMC+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). Finally, strain #36 showed no significant difference between MMC and LB. This suggests that the amount of phages in the environment does not affect the rate of capsule loss, consistent with this strain 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 is inefficient there may be small deleterious impact on the presence of phages able to absorb to the cell. Taken together, these results show that effective predation by induced prophages selects for the loss of the capsule.

DISCUSSION

Temperate phages are major actors of microbial population biology and horizontal gene transfer (HGT). We provide here a first comprehensive analysis of the distribution of prophages in *Klebsiella* genus, their genetic composition and their potential to excise infect and lysogenize other strains. The number of prophages varies very 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 the biology of *Klebsiella*. Prophages differ markedly across strains and their rapid turnover, already observed in other species, may contribute to the phenotypic differences between strains. This is because the association between the number of prophages and genome size (Touchon et al 2016), suggests that bacteria with more lysogens end up acquiring adaptive genes that remain in the chromosome. Since many of these prophages seem to retain the ability to excise, form viable virions, and lysogenize other bacteria, their implication in the biology of *Klebsiella* could be important, either because they may encode adaptive traits or because they may facilitate the transfer of bacterial genes by transduction (including antibiotic resistance in clinical settings). Further work will be needed to assess the phenotypic consequences of lysogeny and the frequency of transducers, 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 molecular mechanisms shaping them (Flores et al 2011, Weitz et al 2013). These analyses use isolated phages from laboratory stocks, or phages recovered from co-evolution experiments (Flores et al 2011). Additionally, these studies focus on virulent phages rather than temperate, either because they envision some sort of phage therapy or because virulent phages give simpler phenotypes. Most studies on prophages are computational and describe the different phage families and the beneficial traits, such as virulence factors, they may code (Bobay et al 2014, Brueggemann et al 2017, Castillo et al 2018). Here, we sought to generate a network of the prophages in naturally infected *Klebsiella* spp. This addresses explicitly the question of what kind of interactions one expects between different *Klebsiella* strains when they meet and their

prophages are induced. These results are especially pertinent for *K. pneumoniae* in clinical settings because many antibiotics stimulate prophage induction (Wagner and Waldor 2002). Also, phages in the mammalian gut, the most frequent habitat of *K. pneumoniae*, tend to be temperate and result from *in situ* prophage induction (De Paepe et al 2016). We found that 60% of strains could produce a lysate that infects at least one strain. This is comparable to what is observed in other species like *Pseudomonas aeruginosa* (66%) (Bondy-Denomy et al 2016) and *Salmonella enterica* (68%) (Zhang and LeJeune 2008). However, we only observed 6% of all the possible cross-infections. This is much less than in *P. aeruginosa,* where a set of different lysogen strains derived from PA14 was infected by *ca* 50% of the lysates tested (Bondy-Denomy et al 2016). Similarly, in *S. enterica* ~35% of cross-infections were effective (Zhang and LeJeune 2008). This result 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. Finally, this also implies that phage-mediated HGT in *Klebsiella* may not be very efficient in spreading traits across the species, and could slow down adaptation.

The small number of cross-infectivity events resulted in a very sparse infection matrix, in which there are only 6% of cross infections. We hypothesized that this low number could be explained by either phage limitations, such as a reduced or inefficient adsorption to the cell, or by unsuccessful DNA injection, or bacterial characteristics, like the presence of other prophages in the genome or bacterial immune systems. More particularly, after viral DNA enters the cell, its expression can be repressed by existing prophages or by DNA inactivation, the so-called, resistance to superinfection by other phages (Bondy-Denomy et al 2016). However, our analysis showed that few prophages were sufficiently similar to potentially provide resistance. Also, some prophages were able to re-infect their original hosts, suggesting that superinfection resistance is not completely efficient in some cases (5 out of 35). Alternatively, some phages could code for a defence mechanisms that protects the host from unrelated phages, as recently discovered in mycobacteriophages (Gentile et al 2019), limiting the number of possible infections. Finally, our results also indicate that bacterial defence systems, like CRISPR-Cas or restriction-modification enzymes, do not play a key role in protecting against exogenous phage DNA.

Our results also show that phage-bacteria infections are highly modular. When we classed the strains in terms of serotype, there was a vast over-representation of efficient infections between bacteria of the same serotype. This serotype specificity is in line with a recent study that focuses on a carbapenemase-producing *K. pneumoniae* CG258, in which very few phages could lysate bacteria with different capsule serotypes (Venturini et al 2019). *Klebsiella* phages have been shown to carry depolymerases that digest the oligosaccharidic bonds and lead to capsule degradation (Latka et al 2017). These enzymes are specific towards capsule serotypes (Hsieh et al 2017, Lin et al 2014, Majkowska-Skrobek et al 2018, Niemann et al 1977a, Niemann et al 1977b, Pan et al 2017, Pan et al 2019, Thurow et al 1974) and could thus explain the specificity we observe of certain prophages to capsule serotypes. 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 serotype. The scarcity of depolymerases could be explained by the fact that they are hard to reliably detect bioinformatically due to high sequence divergence as the sequences of the depolymerases described so far are not homologous. Further, most depolymerases have been described in virulent phages. Depolymerases in temperate phages may be different. We could also hypothesize that depolymerases are rarely found in prophages raising the question of the prevalence of such enzymes in both temperate and virulent phages. Finally, the absence of depolymerases in our strains could also imply that there could be other mechanisms underlying capsule-specificity.

Phage predation depends on the interaction of the phage with the capsule serotype. Since this interaction is specific to a serotype and the capsule varies frequently by HGT (Wyres et al 2016), this should result in a very narrow host range for phages in *Klebsiella*. Indeed, we report very few cases (*ca*. 3%) of phages infecting strains with different serotypes. This could even be an overestimation as infections of strain #63 are taken into account and they may not be directly linked to phage activity but to a potential bacteriocin.

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 (Moller et al 2019, Negus et al 2013, Scholl et al 2005). However, in the case of *Klebsiella* prophages, the capsule is a requirement for infection. One possible explanation could be that the capsule may stabilize viral adsorption and allows more time for efficient DNA injection (Bertozzi Silva et al 2016). Phages can increase their ability to infect by binding reversibly to a cell wall or capsule

moieties, that are more exposed and easily accessible by the phage, prior to binding irreversibly to the specific cell receptor (Bertozzi Silva et al 2016). Further, phages may have co-evolved to recognize a given capsule serotype. Our results suggest that phages increase the selection pressure for the loss of capsule, which results in resistance against phage infection. Since most strains are capsulated, we presume that the non-capsulated bacteria rapidly acquired a novel capsule. This would allow bacteria to be infected by different phages and potentially bring new genes by transduction.

Finally, our results might provide insights in the 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 lead to capsule loss. While phage therapy, may be an ineffective in fully clearing *Klebsiella* infections due to the diversity of serotypes, it can select for non-capsulated mutants. The latter are expected to be less virulent (Paczosa and Mecsas 2016), because the capsule is a major virulence factor in *K. pneumoniae*. Further, the emergence of non-capsulated mutants could 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 (Paczosa and Mecsas 2016).

MATERIALS AND METHODS

Strains and growth conditions. We used 35 *Klebsiella* strains that were selected based on based on MLST data, and representative of the phylogenetic and clonal diversity of the *Klebsiella* complex (Blin et al 2017). 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(Blin et al 2017). We corrected wrongful NCBI species annotations using Kleborate typing (Wyres et al 2016). All information about these genomes is presented in Dataset S1.

Identification of prophages. To identify prophages, we used a freely available computational tool, PHASTER(Arndt et al 2016). 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 S4.

Delimitation and functional annotation of prophages. Prophages are delimited by the *attL* and *attR* recombination sites used for phage re-circularization 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. Functional annotation was performed by combining multiple tools: *prokka* v1.14.0(Seemann 2014), pVOG profiles (Grazziotin et al 2017) searched for using *HMMER* v3.2.1 (Eddy 2011), the PHASTER Prophage/Virus DB (version Aug 14, 2019), *BLAST 2.9.0+* (Camacho et al 2009) , and the *ISFinder* webtool (Siguier et al 2006). 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 Viral Quotient (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.

Core genome. i. *Klebsiella spp*. (*N*=255) (Figure 1A and S1). The core genome was inferred as described in (Touchon et al 2014b). 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) (Figure 4). The pangenome was built by clustering all protein sequences with MMseqs2 (v1-c7a89) (Steinegger and Soding 2017) 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 pan-genome 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% (Figure S1).

Capsule serotyping. We used Kaptive, integrated in Kleborate, with default options (Wyres et al 2016).

Bacteriocin detection. We checked for the presence of bacteriocins and other bacterial toxins using BAGEL4 (van Heel et al 2018).

Depolymerase detection. We checked for the presence of 14 different HMM profiles associated with bacteriophage-encoded depolymerases from multiple bacterial species (Pires et al 2016) Table S1). The profiles were matched against the complete set of prophage proteins using HMMERv3.1b2, filtering by the e-value of the best domain (maximum 1e-3)

and the coverage of the profile (minimum 30%). Five additional sequences of depolymerases, validated experimentally in virulent prophages of *Klebsiella* (see references in Table S1), were also matched against our dataset using *blastp*, and filtering by the e-value (maximum 1e-5), identity (40%) and coverage (40%).

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 = 0.2 and either mitomycin C 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 = 0.2. Mitomycin C was added to final 5 µg/mL. In parallel, non-induced cultures were grown. After 4h hours at 37°C, cultures were centrifuged at 4000 rpm and the supernatant was filter through 0.22um. Filtered supernatants was mixed with chilled PEG-NaCl 5X (PEG 8000 20% and 2.5M of NaCl) and mixed through inversion. Virions were allowed to precipitate for 15 min and pelleted by centrifugation 10 min at 13000 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 allow to grow until OD = 0.8. 1mL of bacterial cultured 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. 10 µL of freshly prepared and PEG-precipitated lysates were spotted on the top agar and allowed to grow for 4h 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 allow to grow until OD = 0.8. 250 µL of bacterial cultured 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 24h 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 (Steinegger and Soding 2017) 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 (bbh) 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:

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

where A_i and B_i is the pair *i* of homologous proteins present in *A* and *B* (containing respectively #A and #B proteins), id(A_i,B_i) is the percent sequence identity of their alignment, and min(*A*,*B*) is the number of proteins of the prophage encoding the fewest 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* by allelic exchange. Upstream and downstream sequences of the *wza* gene (> 500pb) were amplified using Phusion Master Mix then joined by overlap PCR. All primers used are listed in Table S4. 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 hours 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 strain, and again into *E. coli* MFD λpir strain (Ferrieres et al 2010), which was used as a donor strain for conjugation into *Klebsiella spp*. Conjugations were performed for 24 hours at 37°. Single cross-over mutants (transconjugants) were selected on Streptomycin plates (200 µg/mL). Plasmid excision was performed after an 24h culture incubated at 25°C and double cross-over mutants were selected on LB without salt plus 5% sucrose at 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.

REFERENCES

Alvarez D, Merino S, Tomas JM, Benedi VJ, Alberti S (2000). Capsular polysaccharide is a major complement resistance factor in lipopolysaccharide O side chain-deficient *Klebsiella pneumoniae* clinical isolates. *Infection and Immunity* **68:** 953-955.

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

Asadulghani M, Ogura Y, Ooka T, Itoh T, Sawaguchi A, Iguchi A *et al* (2009). The defective prophage pool of *Escherichia coli* O157: prophage-prophage interactions potentiate horizontal transfer of virulence determinants. *PLoS Pathog* **5:** e1000408.

Bertozzi Silva J, Storms Z, Sauvageau D (2016). Host receptors for bacteriophage adsorption. *FEMS Microbiol Lett* **363**.

Bessler W, Freund-Molbert E, Knufermann H, Rduolph C, Thurow H, Stirm S (1973). A bacteriophage-induced depolymerase active on *Klebsiella* K11 capsular polysaccharide. *Virology* **56:** 134-151.

Blin C, Passet V, Touchon M, Rocha EPC, Brisse S (2017). Metabolic diversity of the emerging pathogenic lineages of *Klebsiella pneumoniae*. *Environ Microbiol* **19:** 1881-1898.

Bobay LM, Rocha EP, Touchon M (2013). The adaptation of temperate bacteriophages to their host genomes. *Mol Biol Evol* **30:** 737-751.

Bobay LM, Touchon M, Rocha EP (2014). Pervasive domestication of defective prophages by bacteria. *Proc Natl Acad Sci U S A* **111:** 12127-12132.

Bondy-Denomy J, Qian J, Westra ER, Buckling A, Guttman DS, Davidson AR *et al* (2016). Prophages mediate defense against phage infection through diverse mechanisms. *ISME J* **10:** 2854-2866.

Bossi L, Fuentes JA, Mora G, Figueroa-Bossi N (2003). Prophage contribution to bacterial population dynamics. *J Bacteriol* **185:** 6467-6471.

Brisse S, Grimont F, Grimont PAD (2006). The genus *Klebsiella*. *The Prokaryotes*. Springer: New York, USA. pp 159–196.

Brown SP, Le Chat L, De Paepe M, Taddei F (2006). Ecology of microbial invasions: Amplification allows virus carriers to invade more rapidly when rare. *Curr Biol* **16:** 2048-2052.

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

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K *et al* (2009). BLAST+: architecture and applications. *BMC Bioinformatics* **10:** 421.

Campos MA, Vargas MA, Regueiro V, Llompart CM, Alberti S, Bengoechea JA (2004). Capsule polysaccharide mediates bacterial resistance to antimicrobial peptides. *Infection and Immunity* **72:** 7107-7114.

Canchaya C, Proux C, Fournous G, Bruttin A, Brussow H (2003). Prophage genomics. *Microbiol Mol Biol Rev* **67:** 238-276.

Casjens SR, Gilcrease EB, Huang WM, Bunny KL, Pedulla ML, Ford ME *et al* (2004). The pKO2 linear plasmid prophage of *Klebsiella oxytoca*. *J Bacteriol* **186:** 1818-1832.

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

Chen J, Quiles-Puchalt N, Chiang YN, Bacigalupe R, Fillol-Salom A, Chee MSJ *et al* (2018). Genome hypermobility by lateral transduction. *Science* **362:** 207-212.

Chewapreecha C, Harris SR, Croucher NJ, Turner C, Marttinen P, Cheng L *et al* (2014). Dense genomic sampling identifies highways of pneumococcal recombination. *Nat Genet* **46:** 305-309.

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

De Paepe M, Tournier L, Moncaut E, Son O, Langella P, Petit MA (2016). Carriage of lambda Latent Virus Is Costly for Its Bacterial Host due to Frequent Reactivation in Monoxenic Mouse Intestine. *PLoS Genet* **12:** e1005861.

Doorduijn DJ, Rooijakkers SHM, van Schaik W, Bardoel BW (2016). Complement resistance mechanisms of *Klebsiella pneumoniae*. *Immunobiology* **221:** 1102-1109.

Eddy SR (2011). Accelerated Profile HMM Searches. *PLoS Comput Biol* **7:** e1002195.

Favre-Bonte S, Licht TR, Forestier C, Krogfelt KA (1999). *Klebsiella pneumoniae* capsule expression is necessary for colonization of large intestines of streptomycin-treated mice. *Infection and Immunity* **67:** 6152-6156.

Ferrieres L, Hemery G, Nham T, Guerout AM, Mazel D, Beloin C *et al* (2010). 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* **192:** 6418-6427.

Flores CO, Meyer JR, Valverde S, Farr L, Weitz JS (2011). Statistical structure of host-phage interactions. *Proc Natl Acad Sci U S A* **108:** E288-297.

Fortier LC, Sekulovic O (2013). Importance of prophages to evolution and virulence of bacterial pathogens. *Virulence* **4:** 354-365.

Gentile GM, Wetzel KS, Dedrick RM, Montgomery MT, Garlena RA, Jacobs-Sera D *et al* (2019). More Evidence of Collusion: a New Prophage-Mediated Viral Defense System Encoded by Mycobacteriophage Sbash. *MBio* **10:** e00209-00219.

Grazziotin AL, Koonin EV, Kristensen DM (2017). Prokaryotic Virus Orthologous Groups (pVOGs): a resource for comparative genomics and protein family annotation. *Nucleic Acids Res* **45:** D491-D498.

Guy L, Kultima JR, Andersson SG (2010). genoPlotR: comparative gene and genome visualization in R. *Bioinformatics* **26:** 2334-2335.

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

Holt KE, Wertheim H, Zadoks RN, Baker S, Whitehouse CA, Dance D *et al* (2015). Genomic analysis of diversity, population structure, virulence, and antimicrobial resistance in *Klebsiella pneumoniae*, an urgent threat to public health. *Proc Natl Acad Sci U S A* **112:** E3574-3581.

Hsieh PF, Lin HH, Lin TL, Chen YY, Wang JT (2017). Two T7-like Bacteriophages, K5-2 and K5-4, Each Encodes Two Capsule Depolymerases: Isolation and Functional Characterization. *Sci Rep* **7:** 4624.

Julianelle LA (1928). Bacterial Variation in Cultures of Friedlander's Bacillus. *J Exp Med* **47:** 889-902.

Krinos CM, Coyne MJ, Weinacht KG, Tzianabos AO, Kasper DL, Comstock LE (2001). Extensive surface diversity of a commensal microorganism by multiple DNA inversions. *Nature* **414:** 555-558.

Labrie SJ, Samson JE, Moineau S (2010). Bacteriophage resistance mechanisms. *Nat Rev Microbiol* **8:** 317-327.

Latka A, Maciejewska B, Majkowska-Skrobek G, Briers Y, Drulis-Kawa Z (2017). Bacteriophage-encoded virion-associated enzymes to overcome the carbohydrate barriers during the infection process. *Appl Microbiol Biotechnol* **101:** 3103-3119.

Leclercq S, Cordaux R (2011). Do Phages Efficiently Shuttle Transposable Elements among Prokaryotes? *Evolution* **65:** 3327-3331.

Lee CR, Lee JH, Park KS, Jeon JH, Kim YB, Cha CJ *et al* (2017). Antimicrobial Resistance of Hypervirulent *Klebsiella pneumoniae*: Epidemiology, Hypervirulence-Associated Determinants, and Resistance Mechanisms. *Front Cell Infect Microbiol* **7:** 483.

Lin TL, Hsieh PF, Huang YT, Lee WC, Tsai YT, Su PA *et al* (2014). Isolation of a bacteriophage and its depolymerase specific for K1 capsule of *Klebsiella pneumoniae*: implication in typing and treatment. *J Infect Dis* **210:** 1734-1744.

Lu MJ, Henning U (1994). Superinfection exclusion by T-even-type coliphages. *Trends Microbiol* **2:** 137-139.

Majkowska-Skrobek G, Latka A, Berisio R, Squeglia F, Maciejewska B, Briers Y *et al* (2018). Phage-Borne Depolymerases Decrease *Klebsiella pneumoniae* Resistance to Innate Defense Mechanisms. *Front Microbiol* **9:** 00209-00219.

Moller AG, Lindsay JA, Read TD (2019). Determinants of Phage Host Range in *Staphylococcus* Species. *Appl Environ Microbiol* **85**.

Mori M, Ohta M, Agata N, Kido N, Arakawa Y, Ito H *et al* (1989). Identification of species and capsular types of *Klebsiella* clinical isolates, with special reference to *Klebsiella planticola*. *Microbiol Immunol* **33:** 887-895.

Nakayama K, Takashima K, Ishihara H, Shinomiya T, Kageyama M, Kanaya S *et al* (2000). The R-type pyocin of *Pseudomonas aeruginosa* is related to P2 phage, and the F-type is related to lambda phage. *Mol Microbiol* **38:** 213-231.

Nanda AM, Thormann K, Frunzke J (2015). Impact of spontaneous prophage induction on the fitness of bacterial populations and host-microbe interactions. *J Bacteriol* **197:** 410-419.

Navon-Venezia S, Kondratyeva K, Carattoli A (2017). *Klebsiella pneumoniae*: a major worldwide source and shuttle for antibiotic resistance. *FEMS Microbiol Rev* **41:** 252-275.

Negus D, Burton J, Sweed A, Gryko R, Taylor PW (2013). Poly-gamma-(D)-glutamic acid capsule interferes with lytic infection of *Bacillus anthracis* by *B. anthracis*-specific bacteriophages. *Appl Environ Microbiol* **79:** 714-717.

Niemann H, Frank N, Stirm S (1977a). *Klebsiella* serotype-13 capsular polysaccharide: primary structure and depolymerization by a bacteriophage-borne glycanase. *Carbohydr Res* **59:** 165-177.

Niemann H, Kwiatkowski B, Westphal U, Stirm S (1977b). *Klebsiella* serotype 25 capsular polysaccharide: primary structure and depolymerization by a bacteriophage-borne glycanase. *J Bacteriol* **130:** 366-374.

Paczosa MK, Mecsas J (2016). *Klebsiella pneumoniae*: Going on the Offense with a Strong Defense. *Microbiol Mol Biol Rev* **80:** 629-661.

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

Pan YJ, Lin TL, Chen CC, Tsai YT, Cheng YH, Chen YY *et al* (2017). *Klebsiella* Phage PhiK64-1 Encodes Multiple Depolymerases for Multiple Host Capsular Types. *J Virol* **91:** e02457-02416.

Pan YJ, Lin TL, Chen YY, Lai PH, Tsai YT, Hsu CR *et al* (2019). Identification of three podoviruses infecting *Klebsiella* encoding capsule depolymerases that digest specific capsular types. *Microb Biotechnol* **12:** 472-486.

Phanphak S, Georgiades P, Li R, King J, Roberts IS, Waigh TA (2019). 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* **35:** 5635-5646.

Pires DP, Oliveira H, Melo LD, Sillankorva S, Azeredo J (2016). Bacteriophage-encoded depolymerases: their diversity and biotechnological applications. *Appl Microbiol Biotechnol* **100:** 2141-2151.

Randall WA (1939). Colony and Antigenic Variation in Klebsiella pneumoniae Types A, B and C. *J Bacteriol* **38:** 461-477.

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

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

Rocha EP, Danchin A (2002). Base composition bias might result from competition for metabolic resources. *Trends Genet* **18:** 291-294.

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

Samson JE, Magadan AH, Sabri M, Moineau S (2013). Revenge of the phages: defeating bacterial defences. *Nat Rev Microbiol* **11:** 675-687.

Scholl D, Rogers S, Adhya S, Merril CR (2001). 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* **75:** 2509-2515.

Scholl D, Adhya S, Merril C (2005). *Escherichia coli* K1's capsule is a barrier to bacteriophage T7. *Appl Environ Microbiol* **71:** 4872-4874.

Seemann T (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30:** 2068-2069.

Siguier P, Perochon J, Lestrade L, Mahillon J, Chandler M (2006). ISfinder: the reference centre for bacterial insertion sequences. *Nucleic Acids Res* **34:** D32-36.

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

Steinegger M, Soding J (2017). MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat Biotechnol* **35:** 1026-1028.

Susskind MA, Wright A, Botstein D (1974). Superinfection Exclusion by P22 Prophage in Lysogens of Salmonella typhimurium. IV. Genetics and Physiology of sieB Exclusion. *Virology* **62:** 367-384.

Thurow H, Niemann H, Rudolph C, Stirm S (1974). Host capsule depolymerase activity of bacteriophage particles active on *Klebsiella* K20 and K24 strains. *Virology* **58:** 306-309.

Touchon M, Bobay LM, Rocha EP (2014a). The chromosomal accommodation and domestication of mobile genetic elements. *Curr Opin Microbiol* **22:** 22-29.

Touchon M, Cury J, Yoon EJ, Krizova L, Cerqueira GC, Murphy C *et al* (2014b). The genomic diversification of the whole Acinetobacter genus: origins, mechanisms, and consequences. *Genome Biol Evol* **6:** 2866-2882.

Touchon M, Bernheim A, Rocha EP (2016). Genetic and life-history traits associated with the distribution of prophages in bacteria. *ISME J* **10:** 2744-2754.

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

Tzeng YL, Thomas J, Stephens DS (2016). Regulation of capsule in *Neisseria meningitidis*. *Crit Rev Microbiol* **42:** 759-772.

van Heel AJ, de Jong A, Song C, Viel JH, Kok J, Kuipers OP (2018). BAGEL4: a user-friendly web server to thoroughly mine RiPPs and bacteriocins. *Nucleic Acids Res* **46:** W278-W281.

Ventura M, Canchaya C, Pridmore RD, Brussow H (2004). The prophages of *Lactobacillus johnsonii* NCC 533: comparative genomics and transcription analysis. *Virology* **320:** 229-242.

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

Wagner PL, Waldor MK (2002). Bacteriophage control of bacterial virulence. *Infect Immun* **70:** 3985-3993.

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

Winstanley C, Langille MG, Fothergill JL, Kukavica-Ibrulj I, Paradis-Bleau C, Sanschagrin F *et al* (2009). Newly introduced genomic prophage islands are critical determinants of *in vivo*

competitiveness in the Liverpool Epidemic Strain of *Pseudomonas aeruginosa*. *Genome Res* **19:** 12-23.

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

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

Zhang YF, LeJeune JT (2008). Transduction of bla(CMY-2), tet(A), and tet(B) from Salmonella enterica subspecies enterica serovar Heidelberg to S-Typhimurium. *Vet Microbiol* **129:** 418-425.

ACKNOWLEDGEMENTS

We thank Pedro Oliveira for making available the profiles for the RMS systems, and Amandine Perrin for help with building pan-genomes.

FUNDING

M.H. is funded by an ANR JCJC (Agence national de recherche) grant [ANR 18 CE12 0001 01 ENCAPSULATION] awarded to O.R. J.A.M.S. is supported by an EU FP7 PRESTIGE grant [PRESTIGE-2017-1-0012] and an ANR grant [ANR 16 CE12 0029 02 SALMOPROPHAGE] awarded to E.P.C.R. The laboratory is funded by a Laboratoire d'Excellence 'Integrative Biology of Emerging Infectious Diseases' (grant ANR-10-LABX-62-IBEID). The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

COMPETING INTERESTS

Authors declare that we do not have any competing financial interests in relation to the work described.

FIGURE LEGENDS

Figure 1. Prophage distribution in *Klebsiella* genus. A. Rooted phylogenetic tree of *Klebsiella* species used in this study. **B.** Average number of prophages per species. PHASTER prediction for completeness is indicated. Numbers represent the total number of genomes analysed of each species.

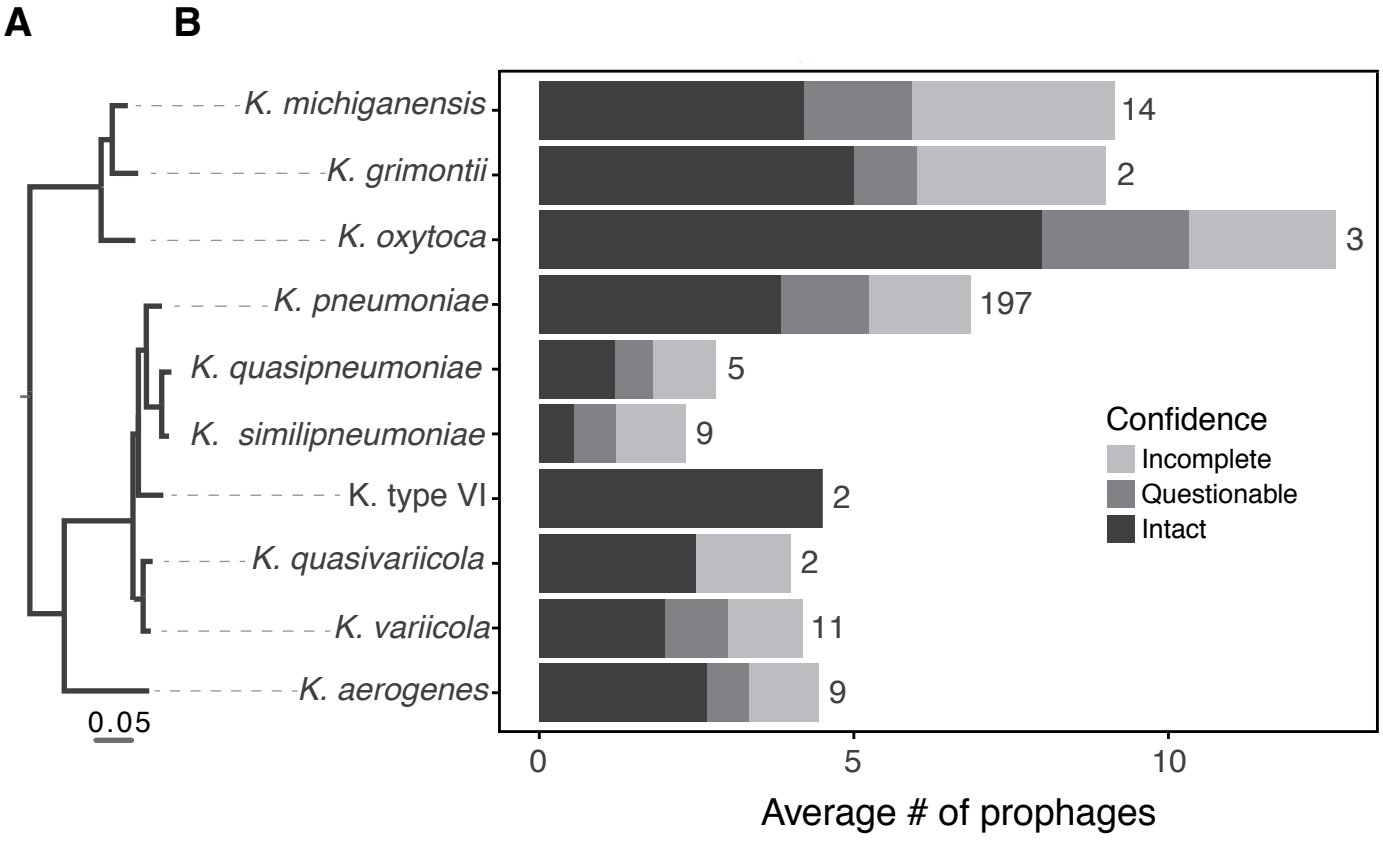
Figure 2. Genomic organization of nine of the prophages in this study. All prophages are classified as "intact" except #63 (1) ("questionable"). Genome boundaries correspond to attL/R sites. Arrows represent predicted ORFs and are oriented according to transcriptional direction. Colors indicate assigned 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 grey meaning that there was no match in the pVOG profiles database and thus no associated VQ value. For prophage #62 (2) that inserted in the core gene *icd*, 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 (Guy et al 2010).

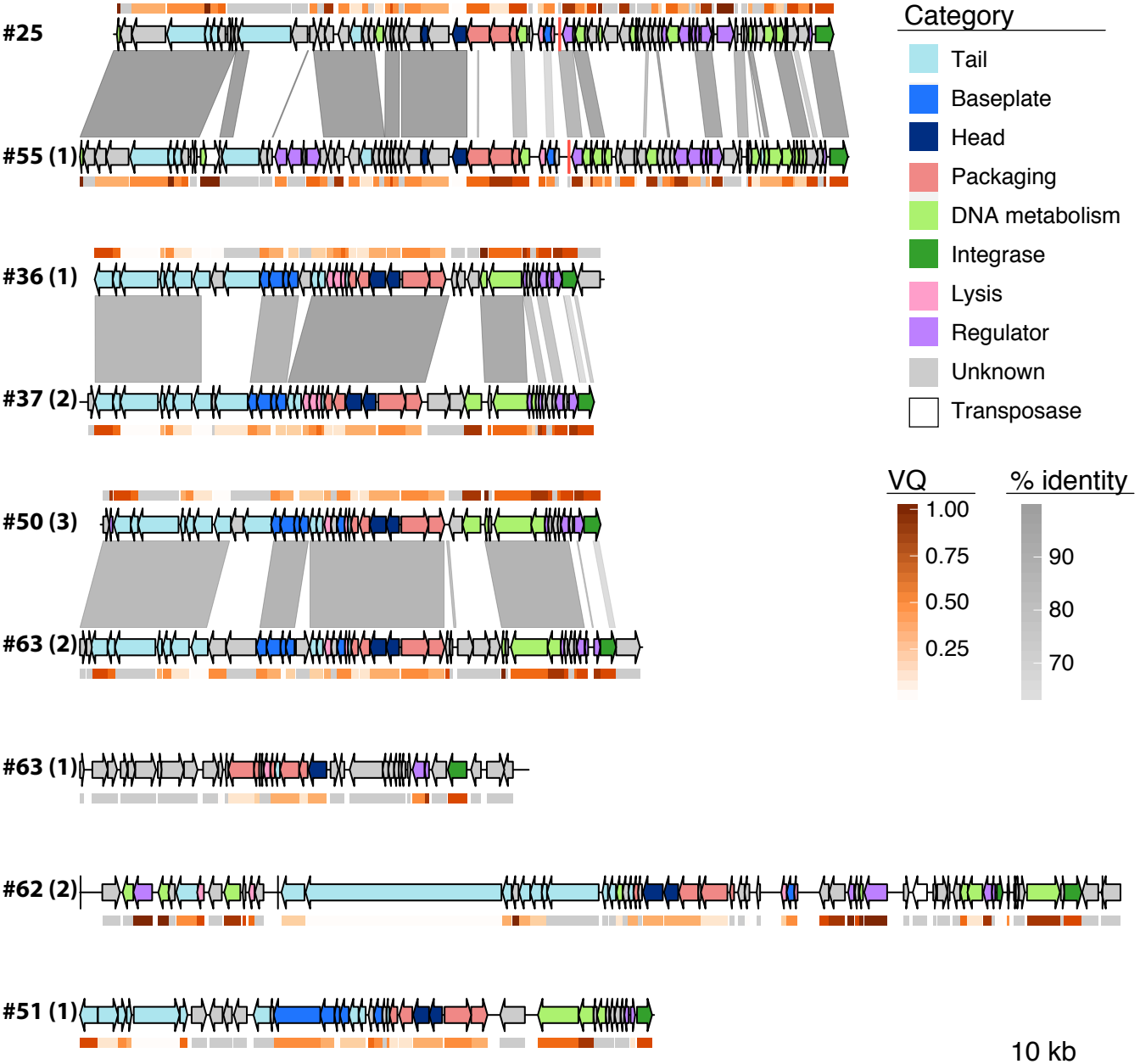
Figure 3. Phylogenetic tree of the 35 *Klebsiella* strains. The tree was built using the protein sequences of 3009 families of the core genome of *Klebsiella spp*. The first column determines the capsule serotype, the second column provides information about the environment from which it was isolated. 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 of mitomycin C. Background colour indicates different *Klebsiella spp*. The size of the circles along the branches are proportional to bootstrap values ranging between 34 and 100.

Figure 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. An average of the three experiments is shown. The strains are ordered by phylogeny, and the colours on top of numbers indicate capsule serotype. Shades of blue indicate infections between lysates produced by bacteria from the same serotype as the target bacteria, whereas shades of orange indicate cross-serotype infections. **B, C and D.** Barplots indicate the number of resistant, sensitive or lysogenized clones from each independent replicate population exposed 24 hours to lysates from other strains.

Figure 5. The capsule plays a major role in phage infection. A. Network of "intact" prophages of 35 *Klebsiella* strains as calculated by the wGRR (indicated in red along the connexions), with a cut-off of 0.5. Node colors represent different serotypes (as in Figure 4). **B** and **C**. Picture of lawns of BJ1(#26) wild type (**B**) and BJ1 capsule null mutant (**C**) in contact with PEG-precipitated supernatant of other strains with KL2 serotype. Isogenic capsule mutant of BJ1 was constructed by an in-frame deletion of *wza* gene (see Methods). **D.** PFU per ml produced by strain #54 after PEG-precipitation of filtered-supernatants. Dashed line indicates the limit of detection of our assay. All experiments were performed in triplicate and error bars correspond to standard deviation of independent biological replicates.

Figure 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 mytomycin 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.




#25
#55 (1)
#36 (1)
#37 (2)
#50 (3)
#63 (2)
#63 (1)
#62 (2)
#51 (1)
Category
Tail
Baseplate
Head
Packaging
DNA metabolism
Integrase
Lysis
Regulator
Unknown
Transposase
VQ
1.00
0.75
0.50
0.25
% identity
90
80
70
10 kb

Isolation
- Environmental
- Community-acquired
- Hospital
- Carriage

Total phages
0, 2, 4, 6, 8

Intact phages
0, 1, 2, 3, 4

Questionable phages
0, 1, 2, 3

Incomplete phages
0, 1, 2, 3, 4

Final OD
0, 0.7, 1.4

Capsule serotype
KL111, KL107, KL2, KL24, KL124, KL1, Other

Capsule, Isolation, Total, Intact, Questionable, Incomplete, Final OD

Kp VI

K. quasipneumoniae subsp. quasipneumoniae
- 08A119 (#39)
- U41 ST256 (#52)
- SB1124 ST2274 (#49)
- 18A069 (#28)
- 01A030T (#36)
- SB98 ST662 (#41)

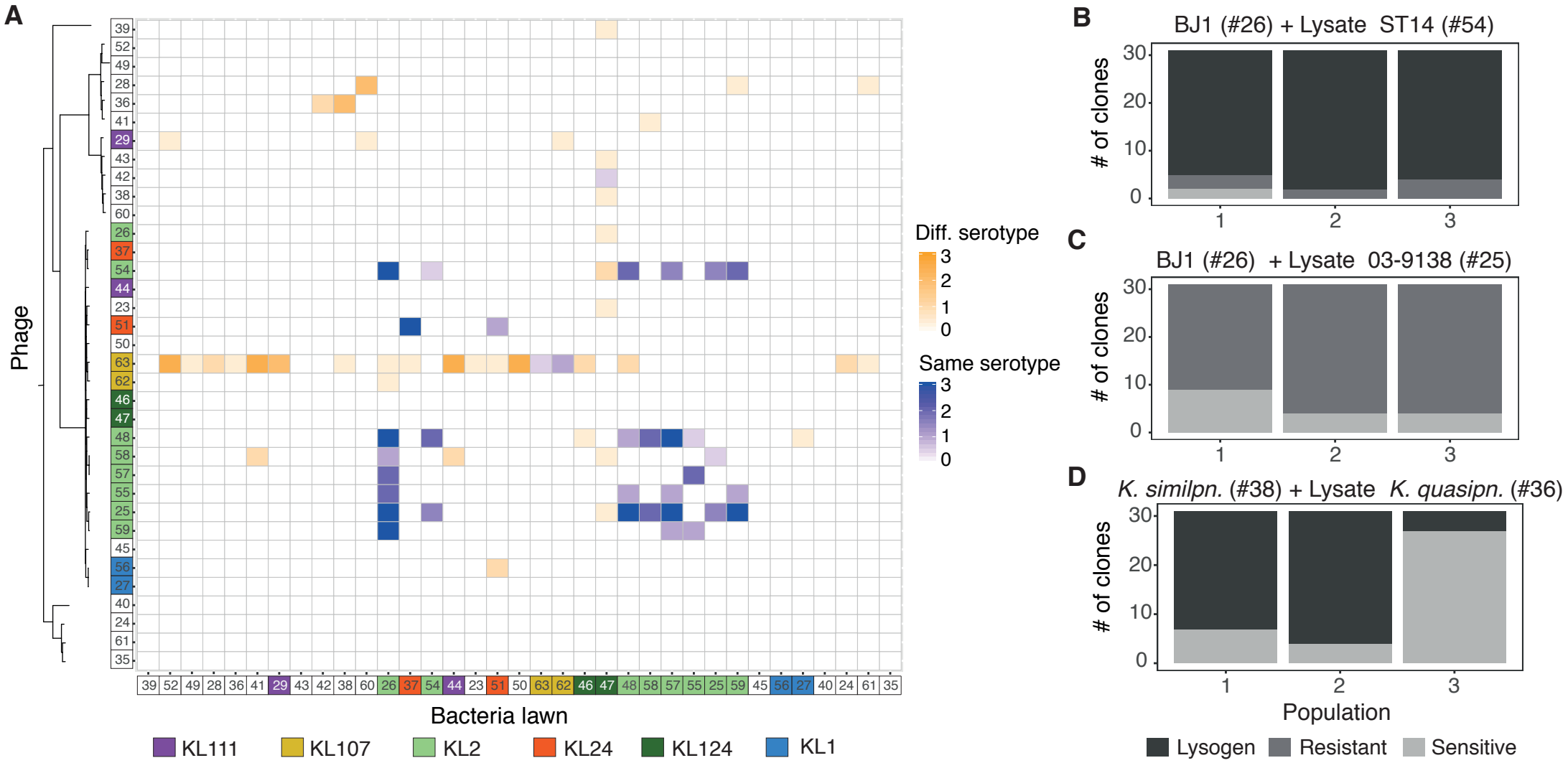
K. quasipneumoniae subsp. similipneumoniae
- SB610 ST2276 (#29)
- 12A476 (#43)
- 09A323 (#42)
- 07A044T (#38)
- CIP 110228 (#60)

K. pneumoniae sensu stricto
- BJ1 (#26)
- 04A025 (#37)
- SB2390 ST14 (#54)
- SB611 ST63 (#44)
- MGH 78578 (#23)
- SB1170 ST45 (#51)
- SB1139 ST37 (#50)
- NJST258-1 (#63)
- Kp13 (#62)
- SB615 ST62 (#46)
- SB617 ST55 (#47)
- SB1067 ST17 (#48)
- CG43 ST86 (#58)
- SA1 (#57)
- CIP 52.145 (#55)
- 03-9138 (#25)
- T69 (#59)
- SB612 ST50 (#45)
- NTUH K2044 (#56)
- SA12 (#27)

Kp V

K. variicola
- CDC 4241-71 (#40)
- 342 (#24)
- At-22 (#61)
- 01A065 (#35)

Tree scale: 0.01


A
Phage
39 52 49 28 36 41 29 43 42 38 60 26 37 54 44 23 51 50 63 62 46 47 48 58 57 55 25 59 45 56 27 40 24 61 35
Bacteria lawn
KL111
KL107
KL2
KL24
KL124
KL1
Diff. serotype
3
2
1
0
Same serotype
3
2
1
0
B
BJ1 (#26) + Lysate ST14 (#54)
# of clones
30
20
10
0
1
2
3
C
BJ1 (#26) + Lysate 03-9138 (#25)
# of clones
30
20
10
0
1
2
3
D
K. similpn. (#38) + Lysate K. quasipn. (#36)
# of clones
30
20
10
0
1
2
3
Population
Lysogen
Resistant
Sensitive