VNIVERSITAT DE VALÈNCIA

The effect of the genetic background on phage infectivity in an encapsulated host

PhD thesis

Doctoral Programme in Biomedicine and Biotechnology

Beatriz Beamud Aranguren

Supervisors:

Rafael Sanjuán Verdeguer

Fernando González Candelas

Pilar Domingo Calap

Valencia, March 2023

Dr. Rafael Sanjuán Verdeguer, Profesor Titular de Genética de la Universitat de València

Dr. Fernando González Candelas, Catedrático de Genética de la Universitat de València

Dra. Pilar Domingo Calap, Investigadora Ramón y Cajal en la Universitat de València

CERTIFICAN: Que Dª Beatriz Beamud Aranguren ha realizado bajo su dirección y supervisión el trabajo que lleva por título "The effect of the genetic background on phage infectivity in an encapsulated host", para optar al Grado de doctora en Biomedicina y Biotecnología por la Universitat de València. La memoria de este trabajo cumple los requisitos científicos y formales para presentar la misma en depósito y proceder a la defensa de la Tesis Doctoral.

Y para que conste, en cumplimiento de la legislación presente, firman el presente informe favorable en Valencia, a 23 de Febrero de 2023.

Dr. Rafael Sanjuán Verdeguer	Dr. Fernando González Candelas	Dra. Pilar Domingo Calap
Co-director	Co-director	Co-directora

A mi padre,

“I have that priceless quality of being curious about life and things which keeps up my enthusiasm”

Charles Chaplin

This thesis has been developed at the Joint Research Unit “Infección y Salud Pública” between the Fundación para la Investigación Sanitaria y Biomédica de la Comunidad Valenciana (FISABIO - Salud Pública) and the Universitat de València (Institute for Integrative Systems Biology, I2SysBio) thanks to a PhD fellowship (FPU16/02139) from the Spanish MCIU. The research was funded by ERC grants 724519 - Vis-a-Vis and 101019724 - EVADER to R.S., project PID2020-118602RB-I00 (Spanish MICINN), project BFU2017-89594R (Spanish MICINN) to F.G.C and project PROMETEO2016-122 (Generalitat Valenciana) to F.G.C and R.S, ESCMID Research Grant 20200063, project PID2020-112835RA-I00 funded by MCIN/AEI /10.13039/501100011033, and project SEJIGENT/2021/014 funded by Conselleria d’Innovació, Universitats, Ciència i Societat Digital (Generalitat Valenciana) to P.D-C.

Agradecimientos/Acknowledgements

"Nothing in Biology Makes Sense Except in the Light of Evolution" es el título de un famoso ensayo por el biólogo evolutivo Theodosius Dobzhansky. Yo diría que nada en la vida tiene sentido si no hay evolución. Y haciendo una analogía con un paisaje adaptativo, para llegar a un pico, también hay que cruzar muchos valles. Esta tesis es el resultado de un largo camino de evolución personal y profesional que no habría sido posible sin todas las personas que me han acompañado en algún tramo de este.

En primer lugar, quiero agradecer a Fernando que como docente me transmitió su pasión por la biología evolutiva hace ya 10 años. Por todas las oportunidades de formación y por tener siempre la puerta del despacho abierta. Gracias Fernando por confiar en mis ideas, y por darme la libertad y los medios posibles para llevarlas a cabo. A Rafa, por acogerme de manera casi fortuita, por el tiempo invertido, y por todo lo que he aprendido en este proceso. A Pilar, por ser un ejemplo de trabajo, dedicación y ambición. A los 3, por vuestra paciencia, por ser críticos, y enseñarme a serlo.

A Manolo Zúñiga, fue un placer hacer una estancia en tu laboratorio y estirarla en el tiempo lo máximo posible. Contigo aprendí que disfrutar la ciencia no depende tanto de la pregunta, si no del modo en qué consigues la respuesta.

A todos mis compañeros de Epimol, FISABIO y Evolución y Salud, por su generosidad y apoyo. Gracias Alma y Leo por enseñarme a dar mis primeros pasos en el laboratorio. A Concha, gracias por su disponibilidad y ayudarme siempre que lo he necesitado. A María, porque con ella empecé con fagos, seguí con recombinación, y acabé con una amiga. Al resto de compañeros. No ha sido fácil con una pandemia de por medio, unos que empezaban y otros que acababan. Aun así, me llevo momentos maravillosos ya sea en forma de sobremesas, escape-rooms, tartific, salidas falleras o partidos de fútbol más/menos disputados. A todos los divulga-teams con los que he disfrutado como una niña pequeña. Especialmente a Marta, Ana, y Mila por acompañarme en la mayoría de aventuras, entre ellas un viaje inolvidable a Croacia y la de mi alter-ego rapera.

No puedo evitar hacer una mención especial a mi fiel compañera de grado, de máster, de doctorado y de vida, Neris. A veces me planteo como puede ser que nos aguantemos durante tantas horas seguidas, independientemente del estrés o decepción que llevemos encima. Dicen que la admiración es la base más fuerte que existe en una relación y tú eres mi ejemplo de ello. Nuestros caminos se separan después de esta etapa, pero no tengo dudas de que se volverán a juntar. No tengo claro en cuál de todos los escenarios científicos/negocios que hemos barajado durante estos años acabaremos. Quizás no volvamos a ser compañeras de estudios ni trabajo, pero sé que contaré contigo para toda la vida. Gracias por acompañarme.

A mi familia valenciana. Gracias Clara y Aída por inspirarme tanto y seguir haciéndolo, por ser mi vía de escape favorita. Gracias Sergio por cuidarme y sacar tiempo y abrazos cuando era necesario. A Irene y Carmen, por quererme y apoyarme tal como soy. A Sandra, por ser uno de los mejores descubrimientos del laboratorio y por el comboi infinito. A los bioinfos y Ángel, por todos los buenos momentos, con discusiones, cervezas, surf, escalada, o lo que se nos ponga por delante. Todos habéis hecho de Valencia mi casa a la que siempre querré volver.

No me puedo olvidar tampoco de Pablo. Gracias por haber confiado en mí siempre y no dejarme rendirme. No hay líneas suficientes en este trabajo para agradecerte, no solo esta tesis, sino también la persona en la que me he convertido.

A mi familia olivareña, de sangre y de calle. Gracias especialmente a María, los Javieres, Soraya y David, por acogerme siempre en su casa y preocuparse por mis batallitas en el laboratorio. A mis dramaqueens. A las de Barcelona. Y a los de siempre.

To my growing Parisian family. It is not been easy to move to another country, to another lab, to a different project, with this work in progress. Sometimes ending things require much more time than that initially planned and you have understood this perfectly. First, thanks David for trusting me. I hope I can return your trust in the near future. Thanks Kochanie, Kozaroume, Spanish mafia and SynBio team (aka, Boulangerie de Synthèse) for navigating with me between these two seas during the last months. 'Calm seas never made a good sailor', but empty ships neither.

Y finalmente, gracias a mi familia. Esa familia que no se elige, pero que yo escogería una y otra vez con los ojos vendados. A mi hermano, por ser mi otra mitad, tan diferentes, pero tan parecidos. A los dos nos han dicho que tenemos muchos pájaros en la cabeza, pero en realidad, nosotros somos los pájaros. Espero que sigas queriendo volar muy alto y que sepas que, aunque ya no estemos en el mismo nido, siempre podrás contar conmigo. A mi madre, por su paciencia, fortaleza, y amor incondicional. Ojalá poder tener algún día el mismo afán de superación con el que te levantas cada mañana. Gracias por seguir cuidándome, aunque a veces no te lo ponga fácil. A mi padre, mi mayor inspiración en el trabajo y en la vida. Gracias por enseñarme el valor del saber y del buen hacer. Por todas las cosas que nos quedaron por vivir y contar, entre ellas esta tesis. Os quiero y os echo de menos. A cada paso que doy, me empujáis para convertirme en la mejor versión de mí misma. Espero poder llegar algún día.

Como diría Freddy Mercury después de leer a Theodosius Dobzhansky,
Evolution must go on.

Resumen/Abstract

Introducción

Los bacteriófagos (fagos), virus que infectan bacterias, son las entidades biológicas más abundantes y diversas del planeta. Como tal, los fagos modulan la evolución y ecología de sus hospedadores y las comunidades en las que estos se encuentran. Por otro lado, la terapia fágica se presenta como una de las alternativas más prometedoras para tratar infecciones causadas por bacterias resistentes a antibióticos. Sin embargo, los mecanismos que gobiernan las interacciones fago-bacteria no se conocen bien, ya que la mayoría de los estudios solo abordan pares individuales fago-bacteria o no tienen en cuenta las características genómicas de ambos. Los avances en secuenciación masiva y metagenómica han permitido vincular los fagos a sus hospedadores a través de firmas genómicas. Aunque muy útiles, estos métodos no proporcionan información individual sobre las interacciones fago-bacteria. Recientes programas bioinformáticos pueden predecir el rango de hospedadores de los fagos con una precisión moderadamente buena a nivel de familia y género, pero obtienen peores resultados a niveles taxonómicos inferiores. Por tanto, el aislamiento de fagos utilizando cepas de interés pueden complementar estos estudios para comprender mejor las interacciones fago-bacteria.

Uno de los factores más importantes para entender el rol de los fagos en la evolución de sus hospedadores y viceversa, es su especificidad. Múltiples factores determinan el rango de hospedador de los fagos. Normalmente, un fago infecta sólo un subconjunto de cepas de una especie bacteriana determinada. Sin embargo, algunos fagos pueden infectar diferentes especies bacterianas o incluso géneros distintos. El reconocimiento y adsorción del virus a los receptores de la superficie bacteriana es el primer paso en la infección y, por tanto, el principal determinante del tropismo de los fagos. Este paso puede implicar a uno o varios receptores. En este último caso, la unión inicial a un receptor primario va seguida de la unión irreversible a un receptor secundario diferente. Los receptores bacterianos pueden ser cualquier molécula expuesta en la superficie, desde carbohidratos hasta proteínas. La mayoría de bacterias presentan una cápsula extracelular compuesta

por polisacáridos, representando un importante factor de virulencia. Las cápsulas protegen a las bacterias de agresiones físicas y del ambiente y son un determinante importante del tropismo de los fagos. El rol de las cápsulas en la infección por fagos puede ir desde proteger a las bacterias al enmascarar los receptores o permitir la infección al ser utilizadas como receptor. Para esto último, algunos fagos codifican dominios depolimerasa en las proteínas de la cola que reconocen y degradan azúcares específicos de la cápsula. Como resultado, los fagos con depolimerasa son específicos de unos pocos tipos capsulares, restringiendo su rango de hospedador.

Klebsiella pneumoniae es una bacteria Gram negativa de la familia *Enterobacteriaceae* y está incluida en el preocupante grupo de patógenos ESKAPE resistentes a los antibióticos. Debido a su importancia clínica, *K. pneumoniae* es uno de los objetivos principales de la terapia con fagos. Uno de los principales focos de diversidad genética de *K. pneumoniae* es el locus capsular, con más de 180 tipos capsulares descritos hasta la fecha. Esta diversidad resulta de mutaciones puntuales pero también de reordenamientos genómicos y recombinación extensiva. Además, *K. pneumoniae* puede inactivar la producción de cápsula, lo que puede ser beneficioso para la bacteria, permitiéndole prosperar en algunos entornos como las células epiteliales o el tracto urinario. Por todas estas razones, *K. pneumoniae* se ha convertido en un modelo para estudiar el efecto de las cápsulas sobre la eficacia biológica en distintos ambientes. Como tal, *K. pneumoniae* también representa un gran modelo para estudiar la coevolución entre cápsulas bacterianas y fagos.

A julio de 2022, se habían secuenciado más de 555 genomas de fagos que infectan *K. pneumoniae*, lo que representa un aumento de ocho veces en comparación con hace cinco años. Sin embargo, la caracterización genómica de estos fagos no ha ido acompañada de su caracterización biológica ya que solo 25% de los fagos han sido publicados con información más allá de su secuencia. La mayoría de los fagos descritos de *Klebsiella* infectan algunos pocos tipos capsulares, pero determinar la especificidad capsular a partir de las secuencias genómicas de los fagos no es posible hasta la fecha. En su lugar, el procedimiento típico implica extensas pruebas de rango de hospedador junto con la expresión y

purificación de las enzimas candidatas. Además, se han descrito algunos fagos sin depolimerasas, pero sus determinantes de rango de hospedador no se conocen.

Una vez los fagos reconocen receptores de la superficie celular, las bacterias han desarrollado múltiples mecanismos de resistencia. Estos mecanismos van desde la modificación de los receptores de superficie, previniendo la adsorción, hasta un número cada vez mayor de sistemas de inmunidad intracelulares. Estos sistemas pueden interferir con el ácido nucleico de los virus como las repeticiones palindrómicas cortas agrupadas y regularmente interespaciadas (CRISPR), los sistemas de restricción-modificación (RM) o pueden conducir a la muerte celular prematura en el caso de los sistemas de infección abortiva. Todos estos mecanismos de defensa complican aún más nuestra comprensión y predicción de las interacciones fago-huésped en la naturaleza pero también para las aplicaciones prácticas de los fagos.

Objetivos

El objetivo de este trabajo es estudiar los determinantes de las interacciones entre *K. pneumoniae* y los fagos que infectan esta bacteria, especialmente en la interacción entre las cápsulas bacterianas y enzimas depolimerasas de los fagos. Para ello, se plantearon los diferentes objetivos específicos. En primer lugar, crear y caracterizar una colección representativa de cepas de *K. pneumoniae* y de los fagos que infectan esta bacteria. En segundo lugar, estudiar los determinantes genéticos que modulan el rango de hospedador de los fagos de *K. pneumoniae* combinando información fenotípica y genómica. En tercer y último lugar, estudiar el rol de la cápsula bacteria en la infección productiva y dinámicas fago-bacteria.

Metodología

Para abordar el primer objetivo, se seleccionaron aislados de *K. pneumoniae* representativos de la diversidad capsular de la especie, se secuenciaron y caracterizaron diversos fagos de *Klebsiella* previamente aislados, y se comparó la diversidad de las cepas bacterianas y los fagos de *Klebsiella* con las bacterias y fagos de las bases de datos. En primer lugar, se realizó el serotipado *in*

silico de más de 1.000 genomas disponibles del programa de vigilancia epidemiológica de *K. pneumoniae* en la Comunitat Valenciana (España). A partir de este serotipado, se seleccionaron 138 cepas que maximizaban tanto el número de tipos capsulares como la diversidad dentro de cada tipo capsular. La colección resultante se caracterizó en más detalle: genes de resistencia y virulencia, tipo de antígeno O, linaje, presencia de receptores de fagos y mecanismos de defensa intracelulares. Se obtuvieron las secuencias representativas de cada complejo clonal de *K. pneumoniae* de la literatura y se obtuvo el genoma compartido para todas las cepas. Este genoma compartido se utilizó para construir una filogenia que permitiera contextualizar las cepas seleccionadas con la diversidad global de *K. pneumoniae*. En segundo lugar, se amplificaron y purificaron mediante ultrafiltración 71 fagos previamente aislados por el grupo de investigación "Environmental and Biomedical Virology (EnBiVir)". Después de la purificación, se realizó la extracción de ADN de los fagos que se secuenciaron mediante Illumina. Una vez se obtuvieron las lecturas de los fagos, se procedió a su ensamblaje *de novo*, para obtener los genomas completos. Los diferentes genomas se compararon entre sí y con genomas de las bases de datos para poder determinar fagos duplicados y su clasificación taxonómica. Cada fago se llamó con la letra de la familia viral, el número del grupo filogenético y una tercera letra para identificar cada especie dentro de cada grupo. Diferentes programas bioinformáticos fueron empleados para determinar si los fagos presentaban un estilo de vida lítico o lisogénico. En paralelo, se estudió su morfología mediante microscopía electrónica de trasmisión y mediante ensayos de calvas. Por último, se realizó un árbol proteómico para poder comparar la diversidad de los fagos incluidos en este estudio con los fagos de *Klebsiella* de las bases de datos.

Para abordar el segundo objetivo, se realizó la matriz de infectividad de todos los pares fago-bacteria posibles. Para ello, se utilizó el ensayo de superposición de agar ("spot test") que consiste en pipetear unos pocos microlitros de cada fago en un césped bacteriano. Se realizaron tres réplicas biológicas y cada interacción se identificó como positiva si al menos dos de las réplicas resultaban positivas. Cada interacción positiva se comprobó en líquido, analizando las curvas de crecimiento bacteriano en presencia de los fagos a partir de los datos de

densidad óptica a 600 nm durante 16 horas. Para aquellos casos en los que no se observó lisis cuando el ensayo en líquido resultó en una interacción positiva, se cuantificó el título viral tras la co-incubación con el hospedador para poder determinar la producción de progenie. Adicionalmente, se puso a punto una nueva metodología para determinar la actividad depolimerasa de los fagos *in vitro*. Este método estuvo basado en cuantificar los restos de polisacáridos tras la exposición de células estacionarias a los fagos. Toda esta información se combinó con la información genética de las bacterias como el tipo capsular, el antígeno O, el clado filogenético o la presencia de receptores proteicos, y con la información de los fagos como su clasificación taxonómica y el tipo de enzima depolimerasa que codificaban. En concreto, se determinó la sensibilidad de cada rasgo genético en predecir las matrices de infección. Para ello, se tomaron pares de referencia aleatorios fago-bacteria y se determinó la probabilidad de que fagos/bacteria con el mismo rasgo genético presentaran el mismo patrón de infección que los pares de referencia. Este proceso se repitió 100 veces para evitar sesgos por la elección del par de referencia y resultó en una tabla de contingencia a partir de la cual se calculó la sensibilidad de cada factor. Por último, estos resultados se complementaron con predicciones a partir de secuencias de las bases de datos haciendo búsquedas de homología de las proteínas de interés mediante blast en profagos.

Para abordar el tercer y último objetivo, se cuantificó la heterogeneidad del fenotipo capsular en los cultivos bacterianos. En concreto, se plaquearon diluciones de cultivos estacionarios para obtener colonias aisladas y se determinó el número de mutantes acapsulares, con morfología traslúcida en lugar de opaca, por 100 unidades formadoras de colonias (UFC). Esta heterogeneidad se relacionó con el espectro de infección de los fagos. A continuación, se aislaron mutantes espontáneos de cápsula estables y el patrón de infección de los fagos en estos mutantes se comparó con el patrón observado en las bacterias parentales encapsuladas. Específicamente, se determinó si los fagos tenían un efecto lítico en las curvas de crecimiento bacteriano cuando la bacteria era acapsular, dada la infección productiva en la bacteria parental encapsulada. Para poder entender el patrón de infectividad observado, se detectaron y analizaron todas las proteínas reconocedoras de receptores de los fagos, en busca de dominios no-depolimerasa.

Para ello, se predijo la estructura terciaria de las proteínas con AlphaFold y se comparó con otras estructuras de las bases de datos. Por último, se modelizaron las curvas de lisis de los fagos para obtener parámetros relevantes de las interacciones fago-bacteria, como el tiempo de lisis, de emergencia de resistencia o coste de la resistencia.

Resultados

Para determinar el rango de hospedador de los fagos de *Klebsiella*, seleccionamos 138 aislados clínicos de *K. pneumoniae* de la región de Valencia que comprendían 59 tipos capsulares, 14 tipos de antígeno O y 76 linajes distintos. Las 138 cepas eran diversas respecto a su susceptibilidad a antibióticos, presencia de profagos y sistemas de defensa como los sistemas CRISPR-Cas y RM. La reconstrucción filogenética con los principales grupos clonales de *K. pneumoniae* mostró que la colección resultante representaba la diversidad global de la especie. Por tanto, la colección de cepas seleccionada era biológicamente relevante para estudiar el rango de hospedador de fagos en esta especie.

Después del análisis genómico, detectamos 46 fagos diferentes de los 70 secuenciados. De los 17 fagos con réplicas biológicas, la mayoría se aislaron con la misma bacteria/tipo capsular (13 y 15, respectivamente) o a partir de la misma muestra ambiental (12). Sin embargo, también se aislaron fagos muy divergentes de la misma ubicación de muestra y huésped bacteriano. Todos los lugares de recogida de fagos estaban muy próximos, lo que podría explicar la presencia de fagos similares en distintos puntos de muestreo. No obstante, la presencia tanto de fagos muy relacionados como de fagos no relacionados en las muestras también refleja un alto nivel de coexistencia de fagos. Los 46 fagos representativos fueron clasificados en 13 grupos (A1-A3, P4, M5, S6, D7, S8-S11, M12 y S13), y 45 especies. Concretamente, los fagos se clasificaron en 5 familias del orden Caudovirales (*Autographiviridae*, *Drexlerviridae*, *Myoviridae*, *Siphoviridae*, y *Podoviridae*) y en una variedad de géneros (*Drulisvirus*, *Przondovirus*, *Webervirus*, *Gamaleyavirus*, *Mydovirus*, *Yonseivirus*, *Jedunavirus*, y *Roufvirus*). Estos fagos comprendían la mayoría de las familias y subfamilias de fagos de *Klebsiella* descritas anteriormente. Todos los fagos excepto uno presentaban valores de

similitud intergenómica <95% con las secuencias de las bases de datos, sugiriendo que representan nuevas especies. Por tanto, este trabajo aporta 44 nuevas especies de fagos. Además, reportamos 11 fagos con valores de similitud genómica con secuencias de bases de datos por debajo del 70%, representando nuevos géneros. Estos resultados muestran la importancia de llevar a cabo muestreos más amplios para descubrir la diversidad de los fagos y, en particular, de los que infectan a *Klebsiella*.

De las 6.319 combinaciones únicas de fago-bacteria evaluadas, solo 124 resultaron en una interacción positiva mediante el "spot test". El rango de hospedador osciló entre 0 y 19 de las 138 cepas (excluida la cepa de aislamiento), con dos patrones diferentes: la mayoría de los fagos (42/46) infectaron una o unas pocas cepas (mediana: 1, media: 1,79; 1-3 tipos capsulares), pero los cuatro fagos de los grupos S8/S9 presentaban un rango de hospedador más amplio (mediana: 11.5, media: 12.25; 4–16 tipos capsulares). El tipo capsular presentó una sensibilidad del 92% para predecir las interacciones positivas de los fagos no-S8/S9. El linaje bacteriano también fue un buen predictor del tropismo de los fagos no-S8/S9 (60%), aunque este factor estaba altamente correlacionado con el tipo capsular en nuestra colección (Cramer's V = 0.9). En cambio, el antígeno O y 45 proteínas adicionales que codifican para posibles receptores fágicos estaban poco correlacionados con el rango de hospedador (sensibilidad < 10%). Nuestros resultados confirman por tanto, que la cápsula bacteriana es el factor principal en determinar el tropismo de los fagos de *Klebsiella,* restringiéndolo a unas pocas cepas de ciertos tipos capsulares.

Para poder explicar el tropismo capsular de los fagos de *Klebsiella*, se llevó a cabo la detección de dominios depolimerasa. Se detectaron dominios depolimerasa en 37 de los 42 fagos no-S8/S9. De los 37 fagos con dominios depolimerasa, 28 fagos presentaban un único dominio, mientras que 8 codificaban 2 y un único fago presentaba 3. Estos dominios se detectaron también en fagos de *Klebsiella* de la literatura, resultando en un total de 163 dominios depolimerasa que agrupaban en 35 tipos diferentes por identidad aminoacídica. Los tipos de depolimerasa explicaban 58% del tropismo capsular de los fagos de *Klebsiella*, mientras que las características taxonómicas de los fagos como el grupo

filogenético presentaban peor sensibilidad con un 18%. Se observó un elevado grado de transferencia genética horizontal de los dominios depolimerasa, ya que los mismos tipos eran compartidos por fagos muy alejados filogenéticamente. Para poder verificar que los tipos de depolimerasa podían predecir el tropismo capsular de otros fagos, se analizaron las depolimerasas de profagos de *Klebsiella* y se determinó el tipo capsular de la bacteria lisógena. Para 18 de los 24 tipos de depolimerasa analizados, se obtuvieron más de 30 coincidencias, indicando que estas depolimerasas también son frecuentes en fagos lisogénicos. Estas coincidencias estaban distribuidas principalmente en el género *Klebsiella* y en *Escherichia*. En concreto, encontramos una predicción estadísticamente significativa del tropismo capsular en 13 tipos de polimerasa (72%). Estas se asociaron a 13 tipos capsulares (KL2, KL3, KL13, KL23, KL25, KL30, KL35, KL47, KL57, KL63, KL64, KL106 y KL140) que corresponden aproximadamente a un tercio (n = 2.563 secuencias) de la abundancia total de tipos capsulares en la colección global de *K. pneumoniae* (https://kleborate.erc.monash.edu/, n = 8.366 secuencias). Estas predicciones abarcaban un gran rango de identidades aminoacídicas lo que sugiere que la presencia/ausencia de los dominios depolimerasa y no mutaciones puntuales predicen el tropismo capsular.

En siguiente lugar, se determinó si el tipo capsular y depolimerasas también reflejaban interacciones productivas entre los fagos y las bacterias. De las 124 combinaciones positivas mediante el “spot-test”, en 94 se observó lisis mediante la comparación de las curvas de crecimiento bacteriano en presencia y ausencia de fagos. De los 30 pares de bacteria-fagos positivos en el ”spot-test” pero negativos en análisis de curvas, solo cinco presentaron evidencia de infección productiva, tras medir la producción de progenie vírica. Por tanto, de las 124 combinaciones positivas iniciales, 99 se clasificaron como virulentas. Considerando la matriz de infecciones productivas, el tipo capsular continuaba siendo el factor que mejor explicaba el patrón de infectividad de los fagos no-S8/S9, pero con una sensibilidad mucho más baja (53% vs. 92%). Para comprender esto mejor, nos centramos en las 25 combinaciones fago-hospedador que no produjeron infección virulenta a pesar de que los fagos produjeron infecciones virulentas en otras cepas. Como era de esperar, los fagos no-S8/S9 degradaron la cápsula en todas las combinaciones

fago-huésped en las que el fago produjo un “spot”, a pesar de ser avirulentos. Estos resultados indican que los “spots” correlacionan bien con el tropismo capsular y depolimerización capsular de los fagos. Además, de las 14/19 combinaciones de los fagos no-S8/S9 mostrando avirulencia, observamos adsorción reversible e irreversible. Las tasas de adsorción fueron cercanas al 100% en muchos casos, lo que indica la presencia de un receptor funcional y señala que mecanismos post-adsorptivos serían responsables de la avirulencia de los fagos. Tras analizar posibles mecanismos post-adsorptivos, descartamos homoinmunidad por profagos y CRISPR, mientras que otros sistemas de defensa como BREX, RexAB, Druantia, Hachiman, y dGTPases estaban más asociadas con cepas resistentes que con sensibles.

Respecto a los fagos de amplio espectro S8/S9, el tipo capsular seguía siendo el factor que mejor explicaba los patrones de infectividad. No obstante, la sensibilidad era mucho peor para estos fagos que para los de espectro reducido, con una probabilidad del 39% para el “spot test”. Por tanto, el tropismo de los fagos S8/S9 no estaba completamente determinado por el tipo capsular. Para comprobar que estos fagos utilizaban mecanismos alternativos a las enzimas depolimerasa, cuantificamos la degradación del polisacárido capsular. En comparación con la fuerte actividad depolimerasa de los fagos no-S8/S9, los fagos S8 mostraron una baja actividad contra la cápsula y el fago S9a no mostró ninguna. Sorprendentemente, la sensibilidad del tipo capsular para los fagos S8/S9 aumentó cuando se consideraron infecciones productivas a un 46%. En estos fagos, los casos de infecciones avirulentas, no mostraron signos de depolimerización ni de adsorción de los fagos. En concreto, los spots parecían ser el resultado de los fagos S8/S9 infectando una población heterogénea de bacterias encapsuladas/acapsulares. Para poder entender mejor los mecanismos de reconocimiento de los fagos S8/S9, llevamos a cabo el análisis de sus proteínas de la cola. Encontramos que los fagos S8 codificaban dos proteínas de reconocimiento de receptores, una de ellas decorada con diferentes dominios de unión a carbohidratos. Por el contrario, el fago S9a codificaba para una esterasa. Esto explicaría que estos fagos reconozcan carbohidratos en algunas cepas lo que podría ayudarles a acceder a receptores secundarios con sus colas flexibles sin

necesidad de actividad catalítica clara como en el caso de dominios depolimerasa en fagos no-S8/S9.

En la tercera parte de este estudio, encontramos que > 90% de cepas bacterianas de la colección presentaban heterogeneidad capsular con una frecuencia media de 8.97 mutantes acapsulares por 100 unidades formadoras de colonias. Esta heterogeneidad afectó la determinación del rango de hospedador de los fagos ya que las bacterias susceptibles a los fagos S8/S9 de amplio espectro presentaron ~ 20 veces más mutantes acapsulares que las bacterias resistentes (*P* = 0.02). Por el contrario, los fagos de espectro reducido mostraron la tendencia contraria. En términos de la estabilidad del fenotipo acapsular, encontramos diferencias significativas entre las bacterias resistentes y sensibles a los fagos especialistas no-S8/S9, siendo la estabilidad de las primeras un 19% mayor. Un total de 223 mutantes acapsulares derivados de 55 cepas bacterianas diferentes fueron testados contra los fagos, cubriendo un 82% de todas las interacciones positivas posibles. Encontramos que en un 70% de los casos, la pérdida de la cápsula hacía resistentes a las bacterias, previamente sensibles. Estos resultados confirman que la cápsula no solo es un determinante pero también es esencial para la infección de los fagos de *Klebsiella*.

La mayoría de casos de infecciones en las bacterias acapsulares fueron resultado de los fagos de amplio espectro, pero los fagos especialistas también eran capaces de infectar en 37 casos. Este patrón no estaba asociado con el grupo filogenético, ya que los fagos capaces de infectar bacterias sin cápsula estaban distribuidos por múltiples grupos. Parte de este patrón podría explicarse por las proteínas de reconocimiento de los fagos, ya que algunos de ellos presentaban dominios con identidad a proteínas involucradas en el reconocimiento de porinas. De ser este el caso, las proteínas quedarían expuestas al suprimir la producción de cápsula y la infección podría iniciarse. El resto de casos podrían venir dados por un diferente efecto pleiotrópico de las mutaciones inactivadoras de la cápsula, que pueden ser específicas de la cepa bacteriana. Por ejemplo, encontramos que 7 de 21 (33%) de los fagos que infectaban más de una cepa bacteriana mostraban tanto un comportamiento capsular-obligado (ningún mutante acapsular infectado de esa cepa salvaje) como facultativo (al menos un mutante acapsular infectado de esa

cepa) dependiendo de la cepa parental en cuestión. Además, encontramos 18 casos en los que los fagos podrían infectar alguno pero no todos los mutantes acapsulares provenientes de la misma cepa salvaje. Todo esto sugiere que diferentes mecanismos de inactivación de la cápsula, pueden tener diferentes efectos en la infección por los fagos. Por último, se cuantificó la dinámica fago-bacteria a partir de los datos de densidad óptica usando como control el crecimiento de la bacteria en ausencia del virus. Estas dinámicas se relacionaron con el valor de esencialidad de la cápsula para cada fago en una bacteria dada, calculado como la proporción de mutantes resistentes acapsulares entre el total de mutantes acapsulares testados. Encontramos una correlación negativa entre el porcentaje de inhibición de los fagos en el crecimiento bacteriano y la dependencia de estos de la cápsula para la infección (Spearman R=-0.36, $P < 0.01$). Esta correlación negativa venía principalmente dada por los parámetros que nos informaban de la emergencia y coste de la resistencia y no por parámetros de la virulencia de los fagos.

Conclusiones

Las interacciones fago-bacteria modulan múltiples procesos ecológicos que son relevantes para explicar la diversidad y evolución de estos microorganismos, pero también para entender su rol en la salud y enfermedades humanas. El objetivo de este trabajo era estudiar los determinantes de las interacciones fago-bacteria en *K. pneumoniae*, con énfasis en el rol de las cápsulas bacterianas y las enzimas depolimerasas de los fagos. Las conclusiones de este trabajo son las siguientes:

1. La colección de fagos y bacterias resultante de este trabajo es representativa de la diversidad de *K. pneumoniae* y sus fagos para poder estudiar las interacciones fago-bacteria.
2. La diversidad de tipos capsulares en *Klebsiella* restringe el rango de hospedador de la mayoría de fagos, siendo no solo el principal factor de especificidad pero también un requisito para el proceso de infección.
3. Las enzimas depolimerasas de los fagos pueden predecir el tropismo capsular, que puede ser inferido mediante genómica comparativa o con datos de infecciones pasadas de profagos.

4. La predictibilidad del tropismo capsular está limitada por mecanismos de resistencia post-adsorptivos diferentes a profagos o CRISPR-Cas.
5. Los fagos sin depolimerasa presentan un rango de hospedador más amplio. Esto es resultado de un modo de acción polivalente ya que los fagos S8 presentaban ambas, dependencia capsular e independencia capsular.
6. La frecuencia y contingencia de las mutaciones que inactivan la capsula afectan al rango de infección de los fagos y tanto los fagos de amplio espectro como los especialistas pueden infectar bacterias acapsulares.
7. El rol de la cápsula como receptor obligado o facultativo afecta las dinámicas fago-bacteria en relación con el desarrollo y coste de la resistencia pero no en la virulencia.
8. En general, datos de interacciones fago-bacteria con información tanto fenotípica como genotípica, como los proporcionados en este trabajo, son necesarios para mejorar nuestra comprensión de las interacciones entre estos microorganismos en la naturaleza pero también para las aplicaciones prácticas de los fagos.

Author's Declaration and publications

I declare that this thesis is a presentation of original work:

- Sections 4.1 and 4.2 present the extended results of the manuscript entitled 'Genetic determinants of host tropism in *Klebsiella* phages' published in Cell Reports (https://doi.org/10.1016/j.celrep.2023.112048).
- Section 4.3 presents the results of the manuscript entitled 'Contingency of capsule-inactivating mutations shapes interactions of bacteria and phages', in preparation.

Besides my supervisors, I deeply acknowledge third parties that have contributed to some of the work presented in this thesis. Neris García González helped with the phage sequencing libraries, with the comparison of *K. pneumoniae* isolates with clonal complexes (Figure 4.2), and with computing resources for prophage analyses. Three undergraduate students under my supervision also helped with some experiments. Concretely, Mar Gómez and David Saiz helped with strain collection and *wzi* PCR amplification, and Lourdes Tordera helped with depolymerization assays and isolation of acapsular bacteria. Additionally, Mar Gómez helped with the literature revision of depolymerases. Concha Hueso provided technical assistance with the preparation of reagents and plates. Nuria Jiménez and Loles Catalán provided technical assistance with the MiSeq library for phage sequencing. Alejandro Sanz Carbonell helped with AlphaFold. Centro de Investigación Príncipe Felipe microscopy service performed TEM.

Other related works that have not been included in this thesis are the following:

- Alsaadi, A., **Beamud, B.**, Easwaran, M., Abdelrahman, F., El-Shibiny, A., Alghoribi, M. F., & Domingo-Calap, P. (2021). Learning from mistakes: the role of phages in pandemics. *Frontiers in microbiology*, *12*, 653107.
    - I contributed to literature revision and draft parts of the manuscript
- Domingo-Calap, P., **Beamud, B.**, Mora-Quilis, L., González-Candelas, F., & Sanjuán, R. (2020). Isolation and characterization of two *Klebsiella*

pneumoniae phages encoding divergent depolymerases. *International journal of molecular sciences*, *21*(9), 3160.

- I contributed to all bioinformatics analyses and manuscript preparation.

- Domingo-Calap, P., **Beamud, B**., Vienne, J., González-Candelas, F., & Sanjuán, R. (2020). Isolation of four lytic phages infecting *Klebsiella pneumoniae* K22 clinical isolates from Spain. *International journal of molecular sciences*, *21*(2), 425.
    - I contributed to all bioinformatics analyses and manuscript preparation.

List of abbreviations

Abi	Abortive infection
ANI	Average nucleotide identity
ARD	Arms race dynamics
AUC	Area under the curve
Cap-	Acapsular
CBM	Carbohydrate-Binding Modules
CFU	Colony-forming units
CG	Clonal Group
CLT	Capsular Locus Type
CPS	Capsular polysaccharide
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
DF	Defense system
Dpo	Capsular depolymerase
dPR	Differential probability of resistance
dsDNA	Double-stranded DNA
dsRNA	Double-stranded RNA
EDTA	Ethylenediaminetetraacetic acid
EPS	Extracellular exopolysaccharide
FSD	Fluctuating selection dynamics
HGT	Horizontal Gene Transfer
HMM	Hidden Markov model

ICTV	International Committee on Taxonomy of Viruses
IGS	Intergenomic similarity
LB	Luria Bertani
LPS	Lipopolysaccharide
ML	Machine-learning
MLST	Multi-Locus Sequence Typing
MOI	Multiplicity of infection
NLSAR	Networked Laboratory for Surveillance of Antimicrobial Resistance
OD	Optical Density
OLT	O-antigen locus type
OMVs	Outer membrane vesicles
ORF	Open reading frame
PBS	Phosphate Buffer Saline
PEG	Polyethylene glycol
PFU	Plaque-Forming Units
pIAUC	Percentage of phage inhibition
PICI	Phage-Inducible Chromosomal Island
r/m	Recombination to mutation rate
RBD	Receptor-binding domain
RBP	Receptor-binding protein
RM	Restriction-Modification
SEA-PHAGES	Science Education Alliance-Phage Hunters Advancing Genomics and Evolutionary Science

ssDNA	Single-stranded DNA
ssRNA	Single-stranded RNA
ST	Sequence Type
TA	Toxin-Antitoxin
TEM	Transmission Electron Microscopy
TF	Tail fiber
TM	Template modelling
TNR	True Negative Rate
TPR	True Positive Rate
TSA	Tryptic Soy Agar
TSP	Tail spike protein
WHO	World Health Organization
WT	Wild-type

Table of contents

1. Introduction 1
1.1. Bacteriophages, their relevance and diversity 3
1.1.1. Phage discovery and ecology 3
1.1.2 Phage morphology 3
1.1.3 Phage life cycles 6
1.1.4 Phage genomics and classification 8
1.1.5 Phage evolution and HGT 12
1.2 Host-range of phages 14
1.2.1 Defining host-range 14
1.2.2 Determinants of phage adsorption 16
1.2.3 Determinants of productive infections 19
1.2.4 Evolutionary and ecological consequences 21
1.2.5 Studying phage-bacteria interactions 24
1.3. Interplay between capsules and phages: *Klebsiella* as a model 26
1.3.1 The bacterial capsule 26
1.3.2 Phage capsular depolymerases 29
1.3.3 Capsule-phage interactions 32
1.3.4 *Klebsiella pneumoniae* 34
1.3.5 *Klebsiella* phages 36
2. Objectives 39
3. Material & methods 43
3.1 Bacteria and phages 45
3.1.1. Bacteria 45
3.1.2. Phages 45
3.2 Bacteria laboratory methods 46
3.2.1. Culture conditions 46
3.2.2 DNA extraction and *wzi* PCR 46

3.2.3 Sanger *wzi* sequencing 47
3.2.4 Capsule extraction and quantification 47
3.2.5 Isolation of acapsular (Cap-) bacteria 48
3.3 Genome analyses of bacteria 49
3.3.1. De novo assembly and typing 49
3.3.2. Phylogenetic analyses 49
3.3.3 Identification of phage receptors 50
3.3.4 Identification of phage defense systems 50
3.4 Phage laboratory methods 51
3.4.1 Phage titration 51
3.4.2 Phage amplification and purification 52
3.4.3 Phage morphology 52
3.4.4 Phage sequencing 53
3.5 Genome analyses of phages 54
3.5.1 *De novo* assembly 54
3.5.2 Clustering & classification 55
3.5.3 Functional annotation 55
3.5.4 Lifestyle prediction 56
3.5.5 Identification of phage RBPs 57
3.5.6 Classification of RBPs systems 57
3.5.7 Identification and clustering of receptor binding domains 58
3.6 Phage-bacteria methods 60
3.6.1 Spotting assay 60
3.6.2 Bactericidal assay 61
3.6.3 Dot-plaque assay 62
3.6.4 Progeny and adsorption assays 62
3.6.5 Depolymerization assay 63
3.7 Additional bioinformatic and statistical approaches 64
3.7.1 True positive rate (TPR) 64
3.7.2. Cross-validation of RBDs with prophages 65
3.7.3. Mechanisms of avirulent infections 66

3.7.4 Analysis of growth curves 67
4. Results 69
4.1 Kpn-Phage, a collection to study phage-bacteria interactions 71
4.1.1 *K. pneumoniae* collection 71
4.1.2 Phage sequencing 75
4.1.3. Clustering of phages 76
4.1.4 General characteristics of phages 78
4.1.5 Phage classification 83
4.1.6 Contextualization of phages with previously reported *Klebsiella* phages 86
4.2 Genetic determinants of host tropism in *Klebsiella* phages 89
4.2.1 Host tropism of isolated *Klebsiella* phages 89
4.2.2 Quantifying the importance of bacterial traits shaping the spotting matrix 91
4.2.3 Sequence-based prediction of phage capsular tropism 95
4.2.4 Validation of sequence-based phage tropism using prophage sequences 97
4.2.5. Prediction of productive infections 100
4.2.6. Determinants of infection of broad-range phages 102
4.3 Beyond specificity: the role of capsules for *Klebsiella*-phage interactions 107
4.3.1 Capsular heterogeneity and phage host-range 107
4.3.2 The basis of Cap- mutants infectivity 110
4.3.3 Contingency of Cap- infectivity 113
4.3.4 Quantification of phage-bacteria dynamics 116
5. Discussion 121
5.1 The need for bacteria and phage collections 123
5.2 Systematic analysis of *Klebsiella*-phage interactions 126
5.2.1 Predicting the tropism of narrow-range phages 127
5.2.2 From tropism to productive infections 134

5.2.3 The role of capsules in productive infections 137
5.2.4 The pitfalls of host-range prediction for broad-range phages 139
5.3 Phages in natural communities 142
5.4 Applications 144
5.5 Personal reflection and final remarks 146
6. Conclusions 149
7. References 153
8. Supplementary material 189

1. Introduction

1.1. Bacteriophages, their relevance and diversity

1.1.1. Phage discovery and ecology

Bacteriophages, often abbreviated and henceforth referred to as phages, are viruses that infect bacteria. Phages were concomitantly discovered by Frederick Twort and Felix d'Herelle in the early 20th century [1,2]. They are the most abundant and diverse entities on Earth, with estimates of 10^{31} particles [3,4]. Phages can be found whenever susceptible bacteria can grow, from the oceans, soils, and plants, to the human microbiome [5]. As a result, phages affect a myriad of evolutionary and ecological processes. For example, phages contribute to the biogeochemical cycles by releasing nutrients, and affect the diversification and composition of bacterial communities after killing their hosts [6,7]. Phages also have important roles in human health, as they modulate the physiology and metabolism of gut bacteria [8]. They also reduce colonization by opportunistic bacteria [9], and contribute to their virulence and pathogenesis [10]. Finally, phages are present in man-made environments like sewage waters or the industry. The latter is especially relevant in the dairy sector, where phages can affect bacterial fermentations, decreasing productivity [11] (**Figure 1.1**). Given their bactericidal effect, phages were proposed as therapeutic agents soon after their discovery, the so-called **phage therapy** [2]. Today, given the worrisome increase in antibiotic-resistant bacteria, there is a growing interest in using phages to treat bacterial infections [12,13].

1.1.2 Phage morphology

In addition to their ubiquity, phages show a great morphological diversity, with tailed and double-stranded DNA (dsDNA) accounting for the vast majority of described phages [18]. Less known phages have single-stranded DNA (ssDNA), single-stranded RNA (ssRNA), or double-stranded RNA (dsRNA) genomes packaged into tailless virions. Phages also vary in size, ranging from 20-200 nm, but some can exceed 200 nm [19].

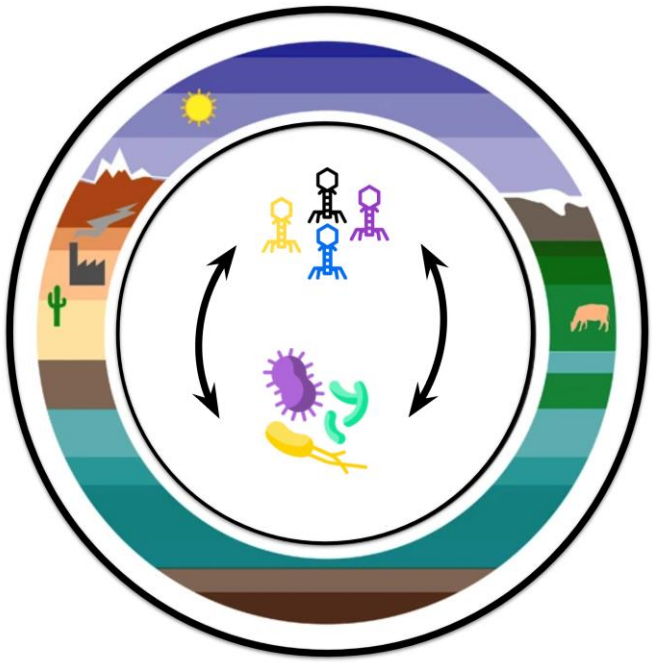

Figure 1.1. Ubiquity and impact of phages in ecosystems. Phages can occur anywhere they can predate bacteria, from aquatic or soil environments to extreme or microbiome-associated ecosystems. Phages are highly abundant in seawater, with estimates of 10^5–10^7 particles per milliliter and a virus-to-bacteria ratio from 1:1 to 100:1 [14]. In soils, phage abundance is variable, being lowest in the desert (10^3 particles per gram) followed by agricultural soils and forests (10^9 particles per gram) [15]. Phages are also found in plants and in the microbiome of animals. For example, the phageome of the human gut is unique for each individual, with estimates of up to 10^8-10^{10} particles per gram of fecal sample [16]. Source: modified from [17].

The traditional classification of phages was based on morphological differences and nucleic acid content (dsDNA, ssDNA, dsRNA, and ssRNA), resulting in six phage types [20]. According to this classification, tailed phages are divided into groups A, B, and C, having long contractile tails, long non-contractile tails, and short non-contractile tails, respectively. Tailless phages are classified into groups D and E, and group F comprises filamentous phages with a rod-like structure (**Figure 1.2**). Phages classified in groups A, B, and C have dsDNA genomes and belong to the 3 commonly used terms *Myoviridae*, *Siphoviridae,* and *Podoviridae*, respectively. These names are still used as different families of the class of tailed phages (*Caudovirales*). Recently, Turner *et al*. have proposed to abolish the order *Caudovirales,* as well as the families *Myoviridae*, *Siphoviridae,*

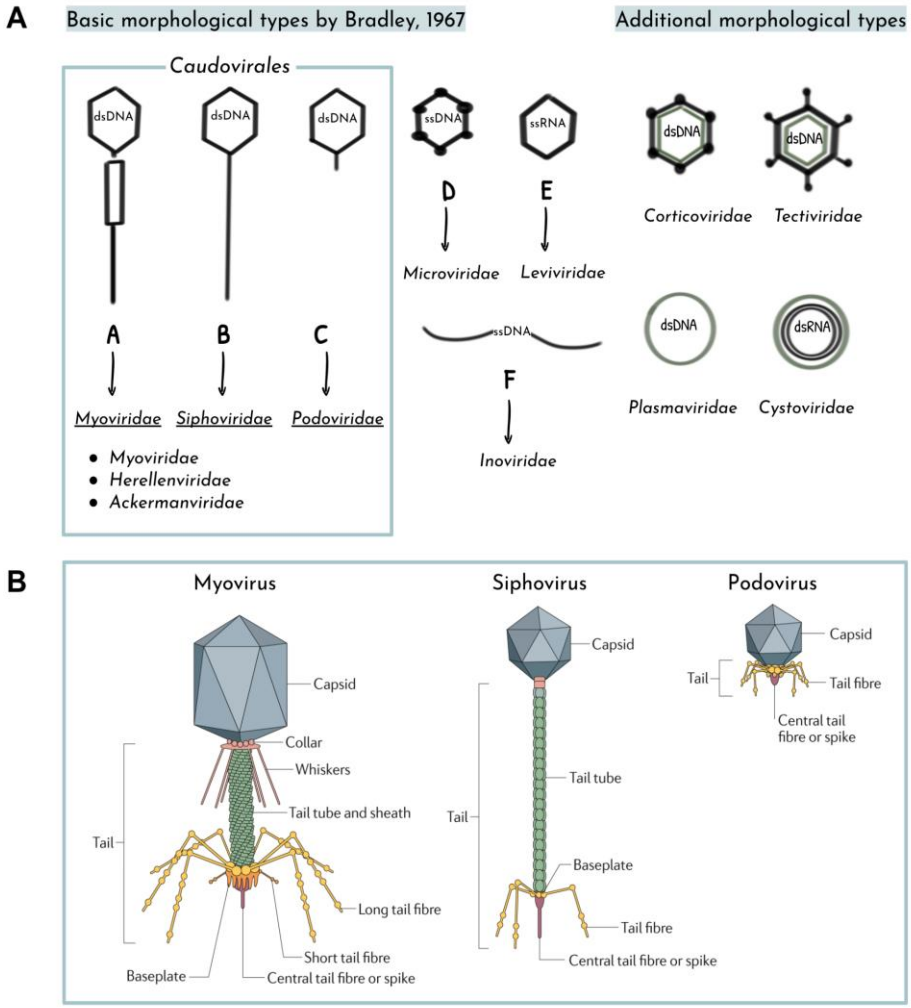

Figure 1.2. Schematic representation of phages morphologies. A. Left, the six morphotypes (A-F) described by Bradley. The current taxonomic names associated with each morphotype by the ICTV are shown below each phage group. The myoviridae morphotype encompasses now the families *Myoviridae*, *Herellenviridae*, and *Ackermanviridae*. Right, additional phage types described after Bradley's classification. These contain lipid layers marked in green. Source: adapted from [5,20]. **B.** Parts of tailed morphotypes. Myoviruses have contractile tails that allow the delivery of phage DNA. In myoviruses and siphoviruses, the tail fibers and spikes are attached to the baseplate. In the case of podoviruses, the tail fibers and spikes are attached directly to the tail. Source: modified from [22].

and *Podoviridae,* due to the lack of evidence for a common ancestor of these taxonomic ranks [21]. Nevertheless, it is still recommended to conserve the terms myovirus, podovirus, and siphovirus to refer to the phage morphology obtained by electron microscopy (morphotypes) (**Figure 1.2**).

Tailed phages possess three main structural components: a capsid, often with icosahedral symmetry, which protects the nucleic acid, a tail that allows the transfer of the phage genome to the host during infection, and an adsorption apparatus at the end of the tail that is involved in host recognition (**Figure 1.2**).

1.1.3 Phage life cycles

The first step in the life cycle of a phage is recognition and adsorption to a suitable host. This process usually involves three phases: initial contact, reversible binding, and irreversible adsorption. Since phages cannot move autonomously, the first phase mainly results from random collisions between phages and bacteria by diffusion. The rate of collisions increases with higher numbers of virions and bacteria, which ultimately increases the adsorption rate [23]. The molecular mechanisms of adsorption are specific for each phage-host combination (**§1.2.2**) and details are only known for some model phages such as *Escherichia* T4 [24]. In general, phage tail fibers/spikes attach reversibly to a primary host surface receptor via non-covalent interactions. This step can help the phage to approach the cell surface for positioning and initiating irreversible adsorption. Binding to the secondary receptor is stronger and renders the phage no longer detachable from the cell [23]. Finally, the conformational change produced after irreversible adsorption allows the ejection of the phage genome into the host [23].

After the genome ejection, the phage can follow **different replication cycles**:

(i) Phages following a lytic cycle hijack the transcription and translation machinery of the bacterium to produce new viral particles. The action of phage lysins leads to rapid cell death (lysis) and release of phage progeny.

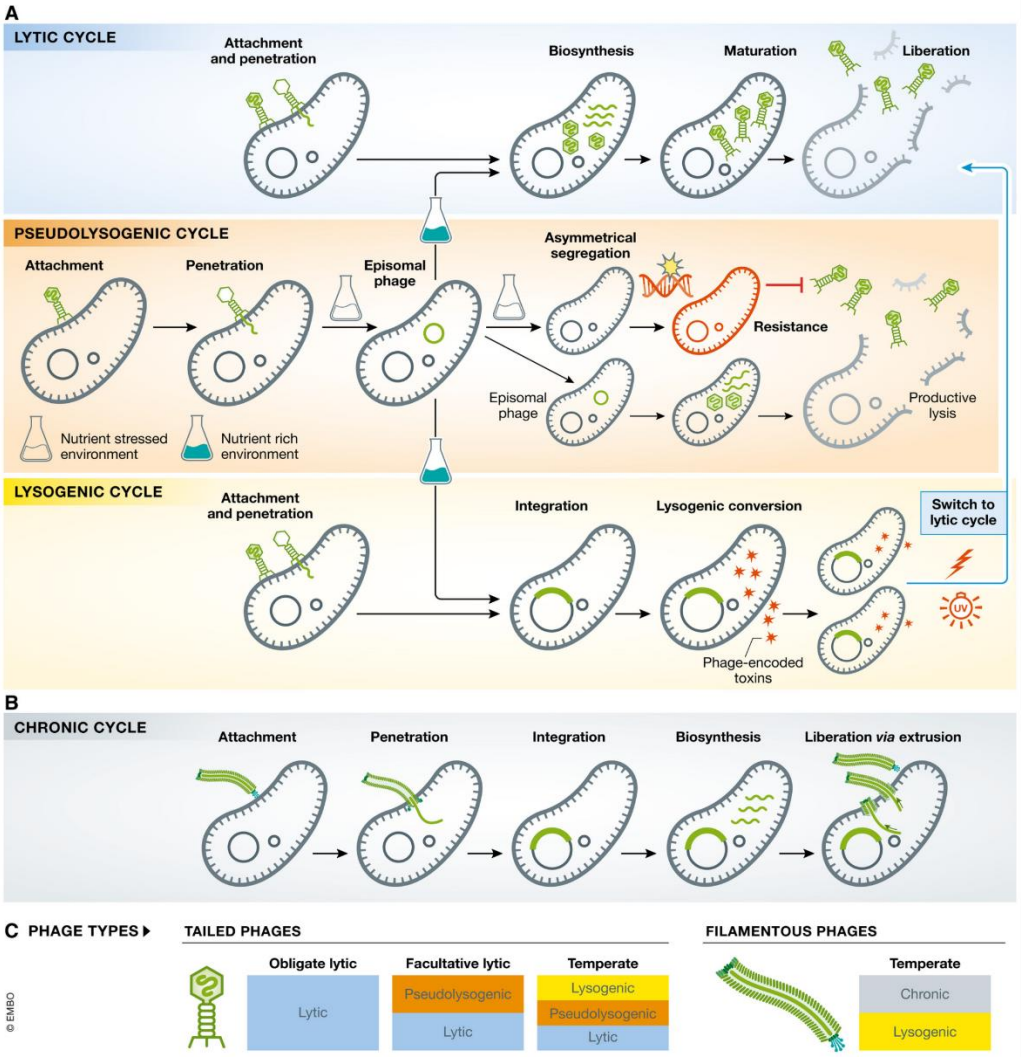

Figure 1.3. Phage lifestyles. A. Life cycles of tailed phages: lytic, lysogenic, or pseudolysogenic. **B.** Life cycle of filamentous phages (*Inoviridae)*: chronic infections. **C.** Classification of phages by their ability to enter the different life cycles. Source: [28].

(ii) Alternatively, phages following a lysogenic cycle can integrate into the bacterial chromosome (becoming a prophage) and remain quiescent upon induction by an environmental signal or spontaneously [25]. Prophages are then inherited concomitantly with the host chromosome that has become a lysogen (Figure 1.3).

(iii) Phages can also follow a pseudolysogenic cycle, in which the phage genome is carried into host cells without stable integration. This can often be observed under low-nutrient conditions and when conditions are favorable, the

virus re-enters a lytic or lysogenic cycle [26]. Meanwhile, the phage remains as an episome in the cytoplasm that is inherited by only one of the descendant cells (**Figure 1.3**).

(iv) Finally, a chronic cycle denotes long-term phage infections in which cells continuously release phage progeny by membrane budding [25] or protein complexes [27] but without lysis.

Based on these four alternatives, phages are classified into obligate lytic, if they only follow the lytic replication cycle, facultative lytic, if they can alternate both lytic and pseudolysogenic replication cycles, and temperate, if they can enter a stable temperate life cycle by integration (**Figure 1.3**). The life cycle of a phage will influence its interaction with the host. For example, lytic phages can act as selective forces for bacteria, favoring the diversification of surface molecules [29] or intracellular defense systems [30]. On the other hand, temperate phages can integrate in the host chromosome and act as switches of bacterial gene expression [31] or drive horizontal gene transfer between bacteria through general, specialized, and lateral transduction [32]. Likewise, pseudolysogeny can promote the coexistence of susceptible and resistant bacteria [33]. This is often referred to as a carrier state, which allows continuous phage production [25] (**Figure 1.3**).

1.1.4 Phage genomics and classification

Phages have a wide range of genome sizes, from the smallest of 2,435 bp (*Leuconostoc* phage L5) [34] to megaphages larger than 500 kb [35]. Genome size is correlated with virion capsid size [36], which imposes a limit on the amount of DNA/RNA which can be encapsidated without destabilizing the virion [37]. The genome of tailed phages encodes at least for DNA packaging and replication functions, the head, the tail apparatus, transcription regulation, and lysis. Phages with larger genomes often encode for more complex functions that could interfere with host metabolism [38].

Although phage classification was historically based on viral morphology, the exponential increase in genome sequences has shifted phage classification to genomic methods. This is the responsibility of the International Committee on the Taxonomy of Viruses (ICTV). For example, the 1999 ICTV report classified tailed phages into 3 families, 16 genera, and 30 species. In 2022, there were 233 families,

168 subfamilies, 2,606 genera, and 10,434 species included (https://talk.ictvonline.org/taxonomy/p/taxonomy_releases). As of July 2022, there were 17,694 complete phage genomes, with 3.7% and 34.5% of them unclassified at the family and genus levels, respectively [39] (**Figure 1.4**). Some of these unclassified groups include phages discovered through metagenomic projects that have been recently isolated or remain to be cultured. For example, a recent study found filamentous phages in microbial genomes and metagenomes, raising the number of *Inoviruses* genomes from 57 to 10,295 [40]. This expansion is also observed for other types of phages, such as huge (>200 kb) [35] or ssRNA phages [35]. Still, dsDNA-tailed phages (*Caudovirales*) ranging from 20 to 200 kb represent more than 75% of the available phage genomes (July 2022, [39]).

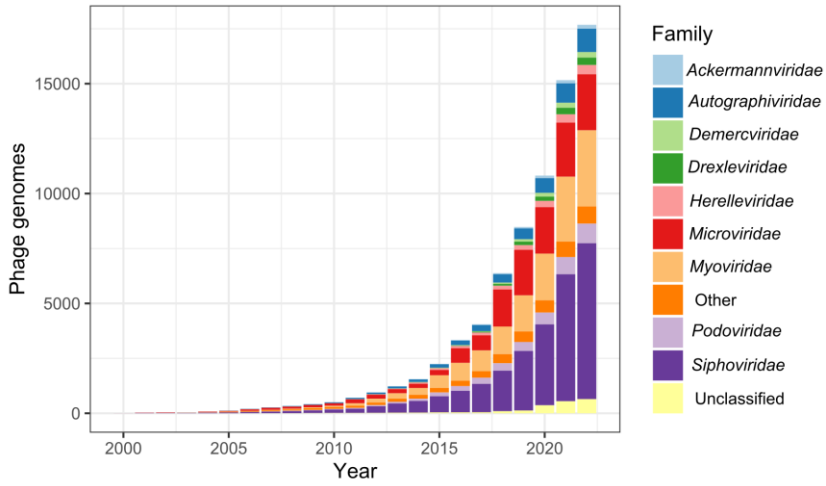

Figure. 1.4. Cumulative number of complete phage genomes in GenBank. Source: This work. Data extracted from https://github.com/RyanCook94/inphared (July 2022).

The overrepresentation of some phages is likely the result of isolation and culture biases, as viruses composed of RNA, with larger sizes or lipid bilayers, are more difficult to detect with current techniques [41]. Overall, metaviromic studies have revealed that some phage groups are more abundant and diverse than previously thought. Moreover, much viral diversity remains to be discovered, the so-called **viral dark matter** [42]. For instance, 60-80% of the open reading frames

(ORFs) in phage genomes are singletons with no similarity to any other ORF [43]. Also, 65-95% of the metaviromes contain novel species [44]. This is also evident if we consider that ~45% of the currently known phages infect only six genera of bacteria (*Mycobacterium*, *Escherichia*, *Pseudomonas*, *Vibrio*, *Streptococcus*, and *Salmonella*) [39]. Biases in bacterial hosts result from the relevance of such bacteria to human health and manufacturing, but also from educational programs of phage hunting. A remarkable example is the SEA-PHAGES (Science Education Alliance-Phage Hunters Advancing Genomics and Evolutionary Science) program, which has isolated a large fraction of available phages, mainly for *Mycobacterium smegmatis* [45,46]. As of July 2022, 2,000 phages against *M. smegmatis* mc^2155 were sequenced as a part of this project (https://phagesdb.org/, accessed 29/07/22).

The increasing knowledge about phage diversity makes genome-based classification very dynamic [21]. Other hurdles to phage classification include the lack of a conserved genetic marker, unlike bacterial 16S ribosomal DNA [47], and the high degree of mosaicism of their genomes [48]. **Genome mosaicism** refers to regions of high sequence similarity, resulting from vertical evolution, with abrupt regions with no detectable identity, resulting from recombination or **Horizontal Gene Transfer** (HGT) from different ancestors [5]. This makes phylogenetic methods, the gold standard for deciphering evolutionary relationships, difficult to apply to phages [49]. In fact, 97% of comparisons between phage pairs result in no detectable DNA similarity [48].

Several strategies have been proposed to study the evolutionary relationships of phages [50]. One of these is the **Phage Proteomic Tree**, which is based on the distances between phage proteomes, and shows high concordance with the ICTV classification [47]. This strategy is especially useful for discerning between higher taxonomic ranks [50]. Also, a network representation of phages is a good alternative to represent phage reticulate evolution [51]. In phage networks, nodes correspond to phage genomes and edges (connections between nodes) represent similarities (**Figure 1.5**). For lower taxonomic ranks, nucleotide identities over complete genomes are preferred. Recently proposed thresholds are 95% and 70% for assigning phages to species and genera, respectively [21]. For this, several

tools can be used, such as VIRIDIC, which calculates **intergenomic similarity** (IGS) values from BLAST comparisons [52], or CD-HIT-EST [53].

In summary, there is no single method for the genomic classification of phages, which is in constant change. A combination of methods is recommended, with priority to those that are independent of functional annotation, or use whole genomes to mitigate the effect of phage mosaicism [50].

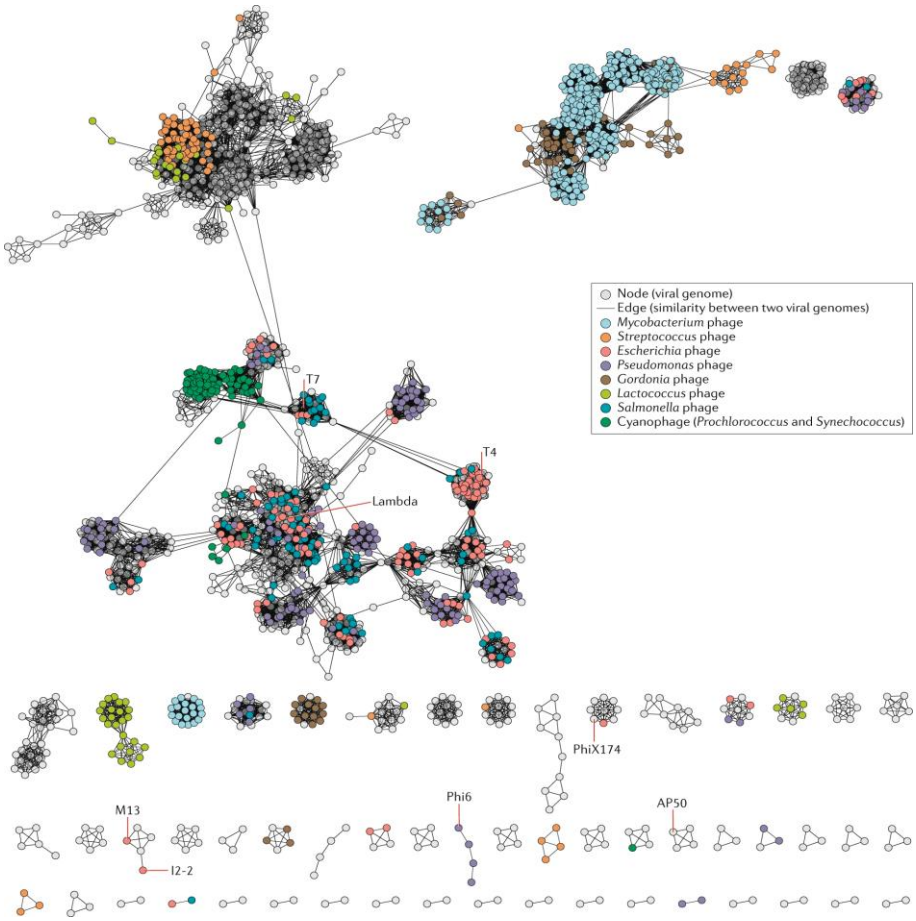

Figure. 1.5. Network representation of phage relatedness. Relationship between RefSeq phage genomes obtained with vConTACT2 (a program for network-based phylogeny) and visualized with Cytoscape v.3.7.1. A node (point) represents each genome and the edges (lines) represent the similarity between pairs of genomes. Phages are colored by host. Source: [5].

1.1.5 Phage evolution and HGT

Phage diversity arises from a high rate of *de novo* mutations, but mostly from the acquisition of genetic material through HGT [54,55]. Estimated mutation rates for bacteria are ~10^{-10} mutations per nucleotide per replication, while mutation rates for dsDNA phages typically exceed these 1,000-fold [56]. This higher mutation rate is due to the error-prone nature of phage polymerases and their lack of proofreading activity [57]. HGT is the exchange of genetic information with phages other than a parent or even with host genomes. This process allows phages to access a vast pool of genes, allowing, for example, their adaptation to new hosts [58,59]. The contribution of recombination compared to mutation (r/m) to phage diversity is high, with estimates of r/m ~ 24 in dairy siphoviruses, much higher than in bacteria [55].

The precise molecular mechanisms underlying the high degree of phage mosaicism are not well understood. The fact that distantly related phages share abrupt regions of high similarity led to propose that HGT was mediated by non-homologous (illegitimate) recombination. During this process, random fragments are end-joined and, if viable phage particles emerge, they can be positively selected [48,60,61]. In other cases, homologous recombination mediated by conserved, short, flanking regions, may account for the exchange of DNA segments [62]. This is supported by the high number of recombinases in phage genomes [63,64].

Recombination or HGT in bacteria and other viruses is often detected by phylogenetic congruence methods [65,66] or by increased density of nucleotide polymorphisms [67]. However, these approaches cannot be applied to phages due to the absence of conserved genes and the high degree of reticulate evolution that can confound all phylogenetic analyses [49,68]. Therefore, recombination is usually determined when phage pairs share a few genes or regions with high sequence similarity. This differs from phages with ancient ancestry, where a high number of genes with low similarity are observed [69](**Figure 1.6**).

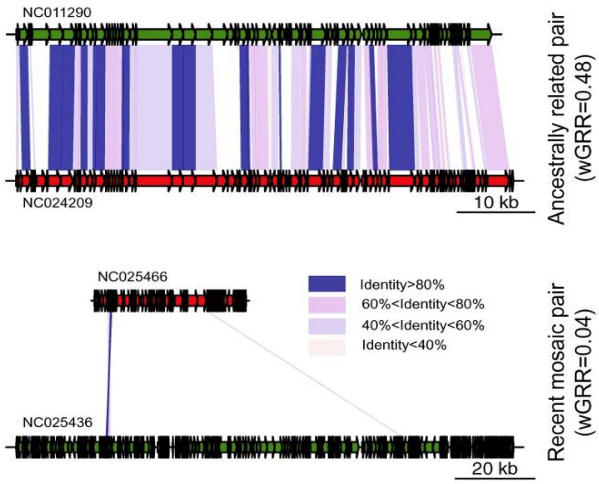

Figure 1.6. HGT in phage genomes. Examples of phage pairs with remote ancestrality (top) or with recent HGT events (bottom). Genes with similarity are connected with the color of lines being proportional to the nucleotide identity. wGRR: weighted Gene Repertoire Relatedness, a measure that integrates the frequency of homologs and the sequence identity. Source: [69].

In general, temperate phages are considered **high gene content flux phages**, as they acquire and lose genes at high rates. On the other hand, virulent phages are mostly classified as **low gene content flux phages** [48,63]. The high flux of temperate phages can be explained by recombination with other temperate phages co-infecting the cell, or between prophages in poly-lysogenic bacteria [63,70]. However, opportunities for virulent phage pairs to recombine are thought to be rare as they lead to bacterial lysis. One explanation might be **cryptic co-infections**, where adsorbed phages could undergo genetic exchanges following DNA cleavage by bacterial intracellular defense mechanisms [54]. Exchanges between lytic and temperate phages were also thought to be rare, as they share very low sequence similarity [48]. Recently, 8% of the total pairs of temperate and virulent phages were found to be mosaics [69].

1.2 Host-range of phages

1.2.1 Defining host-range

One of the key factors in understanding phage diversity and the role of these parasites in the evolution and ecology of bacteria is their host-range. **Phage host-range** is ideally defined as the set of bacteria that phages infect. However, this definition can be misleading as it does not reflect the myriad of interactions between phages and bacteria throughout the complete infection cycle [71,72]. The mechanisms of bacterial resistance are very diverse, ranging from receptor modification, thus affecting the first step of phage adsorption, to an increasing number of intracellular immunity systems [73,74]. Host-range also depends on the conditions and techniques used in its determination [59,75,76]. This led Hyman and Abedon [71] to propose seven categories of host-range:

(i) The adsorptive host-range is the number of hosts on which the phage is adsorbed (**Figure 1.7**). This can be determined by the decrease in the amount of free phage after co-incubation with the host and can be reversible or irreversible (**§1.1.3**).

(ii) The penetrative or transductive host-range denotes the ability of phages to deliver their nucleic acid into the cytoplasm of bacteria after adsorption.

(iii) The bactericidal host-range indicates that phage adsorption results in bacterial death (**Figure 1.7**).

(iv) The plaquing host-range indicates the hosts on which phages can produce plaques. Phage plaques are a clearing in a bacterial lawn resulting from the outward diffusion of phage progeny and are equivalent to bacterial colonies.

(v) The productive host-range is bacteria in which phages can replicate. Phages that produce plaques have productive host-ranges but not forming a plaque is not a definite sign of unsuccessful phage replication. Therefore, the ultimate test for productive host-ranges is to observe an increase in extracellular phage after co-incubation with the host due to progeny release.

(vi) The lysogenic host-range would be the hosts in which phages integrate, forming lysogens. This could be also referred to as **reductive host-range** to encompass also the pseudo-lysogenic state of phages [71].

(vii) The spotting host-range refers to one of the most popular methods for determining phage tropism. **Spots** are areas of bacterial clearance after phage application that can be the result of productive infections. A phage spot can also be caused by "**lysis from without**", which occurs when multiple virions co-adsorb to a single cell and cause sufficient cell wall damage to kill it [77], or by abortive infection, which causes cell death but without phage replication [71]. Recently, we have shown that phage spots can also result from the ability of phages to degrade the bacterial capsule without bacterial killing [59].

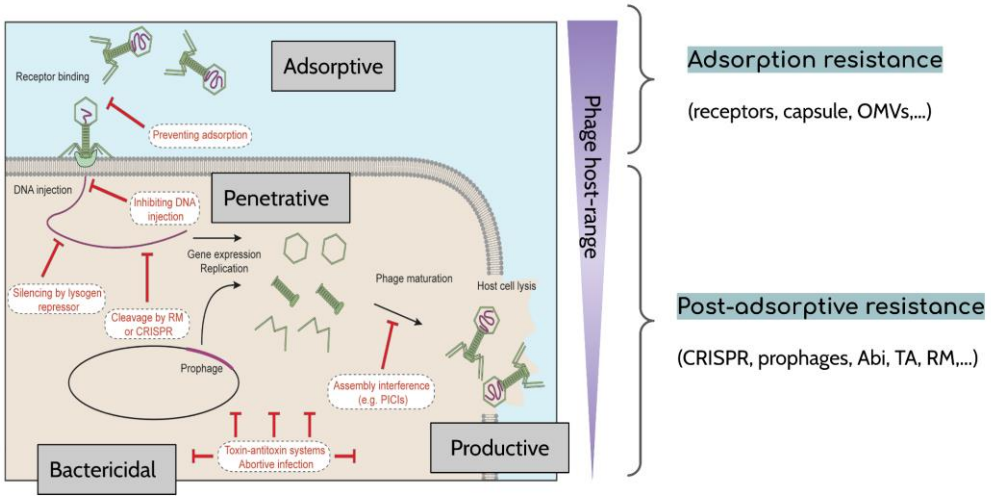

Figure 1.7. Phage host-range and bacteria resistance levels. The productive host range of a phage can be narrowed by different levels of bacterial resistance mechanisms. Phage resistance mechanisms are labeled in red and can generally be divided into prevention of phage adsorption or post-adsorptive resistance. OMVs: Outer membrane vesicles; CRISPR: Clustered Regularly Interspaced Short Palindromic Repeats; Abi: Abortive infection; TA: Toxin-Antitoxin, RM: Restriction-Modification. Source: Modified from [78].

Apart from methodological caveats, each category of phage host-range is the result of the specific interactions between the virus and the host, and their arsenals of infection and resistance mechanisms, respectively.

1.2.2 Determinants of phage adsorption

Recognition of and adsorption to bacterial surface receptors by the virus is the first step in phage infection and, therefore, the primary determinant of phage host-range. Bacterial receptors are surface structures that phages employ for attachment and are often crucial for the metabolism and survival of bacteria. Cell wall composition varies not only between species but also within species, which explains why phages are so specific to a few strains [79]. In Gram- bacteria, the outer membrane is the most exposed layer of the cell and different sugar, proteinaceous and lipidic structures are anchored to it. The following structures can be arranged from the most to the least exposed: the **extracellular exopolysaccharide** (EPS), **capsular polysaccharide** (CPS), **lipopolysaccharide** (LPS, composed of the lipid A, core LPS, and the O-antigen), and outer membrane proteins, such as porins (OmpC, OmpF, TolC, etc.) (**Figure 1.8**). Gram- *Siphoviridae* phages primarily target outer membrane proteins, while *Podoviridae* and *Myoviridae* phages more frequently bind to CPS or LPS polysaccharides [80]. Transient, leaky binding to surface glycans is usually followed by irreversible attachment to a secondary protein receptor [79]. Less frequently, the primary and the secondary receptor are the same structure. Phages can also target bacterial appendages, such as flagella and pili, that serve as primary receptors [81]. Interestingly, the interaction of phages with the primary receptor may or may not be necessary for successful adsorption [82].

The tail machinery of phages often mediates receptor attachment. The proteins involved in these processes are called **receptor-binding proteins** (RBPs) and are usually located at the distal end of the tail. RBPs can be classified into **tail fibers** (TFs), which are long, thin, fibrous proteins that lack enzymatic activity, and **tail spike proteins** (TSPs), which are shorter and thicker (**Figure 1.2**) and often have enzymatic activity [83]. In addition, viral capsids may also mediate phage-host interactions [84]. Enzymatic domains, contained in TSPs, could aid phage adsorption to CPS, LPS, or EPS by specifically cleaving the polysaccharides that make up these layers. They are commonly referred to as hydrolases, if they degrade the peptidoglycan or the wall teichoic acids, or **depolymerases**, if they have activity against LPS or CPS (**§1.3.2**) (**Figure 1.8**).

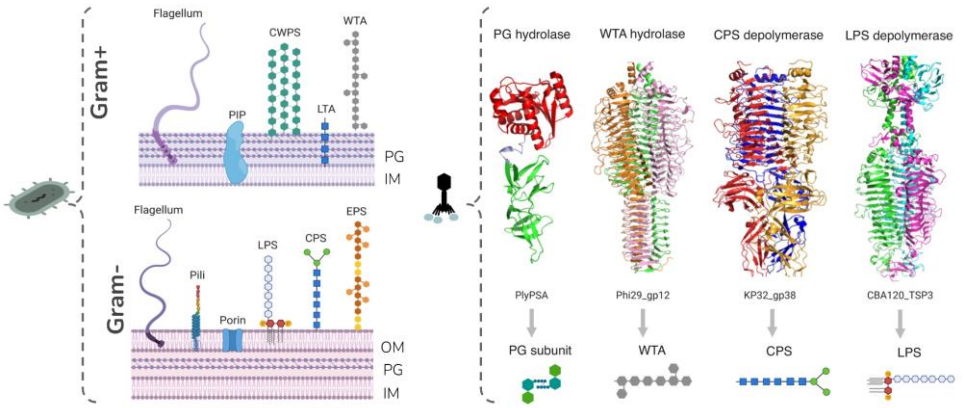

Figure 1.8. Bacterial receptors and phage receptor binding proteins (RBPs). Left: Main bacterial receptors in Gram+ and Gram- bacteria. Some bacteria can be flagellated, presenting pili, or encapsulated, providing additional structures for phage attachment. OM: Outer membrane; PG: Peptidoglycan; IM: Inner membrane; PIP: Phage infection protein; CWPS: Cell wall polysaccharide; LTA: lipoteichoic acids; WTA: wall teichoic acids; LPS: Lipopolysaccharide; CPS: Capsular polysaccharide; EPS: exopolysaccharide. Right: Phage RBPs with enzymatic activity. Top: protein structure of the RBP. Bottom: Substrate of each enzyme. Source: modified from [85].

Bacterial receptors are unknown for most phages, which makes it difficult to decipher phage host-range, and whether infection strategies are shared by other viruses [86]. Additionally, some phages can recognize multiple receptors by acquiring additional RBPs or mutating their existing RBPs. Therefore, we can distinguish between monovalent phages, which target a single bacterial receptor and are more likely to have a narrow host-range, and polyvalent phages, which can bind to several receptors and usually have broader host-ranges [87]. Still, monovalent phages can also present a broad host-range if they target conserved structures (e.g., *Salmonella* phages that bind the LPS core rather than the variable region) [88]. In addition, monovalent phages can become polyvalent. For example, point mutations in the RBP of *Escherichia* phage lambda allow this phage to recognize not only LamB but also OmpF protein receptors [89]. Lastly, polyvalence can result from phages having several RBPs that may or may not be expressed simultaneously. The latter can produce phage quasispecies where each particle can infect one type of host but the population as a whole has a broader host-range [87].

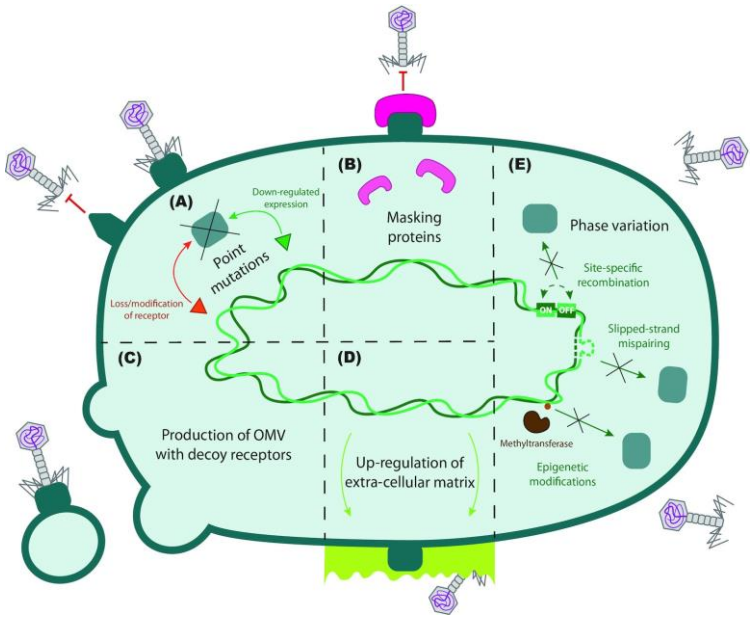

Figure 1.9 Bacterial mechanisms to prevent phage adsorption. (A) Modification of phage receptors (green rectangles) or reduction of their expression by point mutations. **(B)** Production of receptor masking proteins (in pink) that compete for receptor binding with phages. **(C)** Production of outer-membrane vesicles (OMVs) in response to phage infection, thereby decreasing the number of phages available to infect cell surface-available receptors **(D)** Increased production of extracellular matrix (light green) to mask phage receptors. **(E)** Phase variation of bacterial receptors results in phenotypic resistance to phages. Source: [90].

Just as phages can adapt to new hosts, bacteria can develop resistance mechanisms to avoid phage adsorption. In general, these fall into two categories: receptor loss and phage-encounter block [90]. Firstly, bacteria can lose, modify, or down-regulate phage receptor expression. For example, *Acinetobacter baumannii* becomes resistant to phages øFG02 and øCO01 by mutating genes involved in the CPS synthesis [91]. In addition to point mutations, **phase variation** can also affect phage resistance through reversible loss or down-regulation of phage receptors [92,93](Figure 1.9). Secondly, bacteria can prevent phage access to their receptor by increasing EPS or CPS production [94], or by secreting masking proteins that compete with phage binding. Finally, the production of outer membrane vesicles (OMVs) by bacteria can also prevent phage union to the cell surface (Figure 1.9).

1.2.3 Determinants of productive infections

In case phage adsorption is successful, the virus still has to circumvent several intracellular bacterial defenses. Post-adsorptive resistance mechanisms can be broadly categorized into four categories [78]:

(i) Prevention of DNA injection: **superinfection exclusion** of residing prophages can prevent incoming phages to inject their DNA. For example, the *Escherichia* prophage HK97 prevents infection against other HK97 or related HK75 phages by expressing a transmembrane protein that prevents translocation of the viral genome across the inner membrane [95].

(ii) Nucleic acid interference: prophages can also interfere with the DNA of invading phages by expressing their cognate transcriptional repressor. This DNA-binding protein has the ability to suppress the integration of new related phages or the expression of genes relevant to their lytic cycle [96,97]. Also, restriction-modification (RM) and clustered regularly interspaced short palindromic repeats and associated proteins (CRISPR-Cas) systems can degrade viral DNA through the action of nucleases (**Figure 1.10**). **RM systems**, which consist of a restriction endonuclease and a methyltransferase, are ubiquitous anti-phage systems [98]. The endonuclease recognizes 4-8 bp motifs and cuts the viral DNA. These motifs are also present in the host, which protects its own DNA by methylation via the system's methyltransferase [99]. **CRISPR-Cas systems** are present in nearly half of the sequenced bacteria and are also highly diverse [100]. So far, two classes, six types, and 33 subtypes of CRISPR-Cas systems have been described [101]. They are composed of a CRISPR array and a *cas* gene operon that encodes for Cas proteins. Roughly, all CRISPR-Cas systems have the same mechanism. First, the invading DNA is excised into short fragments (30-40 bp), called spacers, and incorporated by Cas proteins into the CRISPR array flanked by short semipalindromic repeats. Second, the spacers and repeats are transcribed and processed into CRISPR RNA guides (crRNA). Third, Cas proteins form a complex with the crRNA guide. This complex can recognize and bind a complementary sequence to the crRNA, which leads the Cas nuclease to cleave the new invading DNA [78]. This system represents an adaptive immune mechanism of bacteria against phages that can create a 'memory' similar to that of the eukaryotic immune

system, but inherited by the bacterial progeny [102]. Other recently described systems such as **DISARM**, **BREX**, or **Ago** also interfere with phage DNA (**Figure 1.10**), preventing its replication [90].

(iii) Abortive infection (Abi): This process consists of 'programmed' cell death by bacteria that prevents phage virions from spreading to nearby cells. This is achieved by different processes that can be triggered by various systems [103]. For example, an unbalanced ratio of **toxins and antitoxins** molecules in the cell could lead to Abi [104]. In addition, the action of effector proteins, in the case of **Retrons**, or secondary messengers, like in the **CBASS** or **Thoeris** systems (**Figure 1.10**), could lead to membrane degradation or NAD depletion, processes that ultimately result in programmed cell death [90].

(iv) Assembly interference: **Phage-Inducible Chromosomal Islands** (PICIs) parasite phages for their own replication by repressing the expression of phage late genes. As a result, the genomes of the PICIs, rather than the genome of the infecting phage, are preferentially encapsidated [105].

All these intracellular defense mechanisms are probably just a snapshot of all the strategies that bacteria have evolved to overcome phages during their coevolution. This has become clear in recent years, in which an increasing number of defense systems, whose mechanisms are still unknown, have been described. These include Hachiman, Shedu, Gabija, Septu, Lamassu, Zorya, Kiwa, Druantia, Wadjet, AVAST, and Paris, among others [90].

Phages can also possess counter-defenses against the bacterial intracellular mechanisms mentioned above, which will impact their host-range. For instance, some phages lose motifs recognized by RM enzymes [106], modify their DNA using non-canonical nucleotides [107], or encode methyltransferases to methylate their own DNA and protect it from degrading [108]. Certain phages carry antiCRISPR proteins that inhibit the binding of the cRNA-Cas complex to target DNA [109]. Phages can also produce their own antitoxin to prevent abortive infection triggered by bacterial toxin-antitoxin systems [110]. Finally, phages should be able to counteract recently discovered systems. For example, the Ocr protein of phage T7, which is able to evade RM systems, was also found to be active against the BREX system [111].

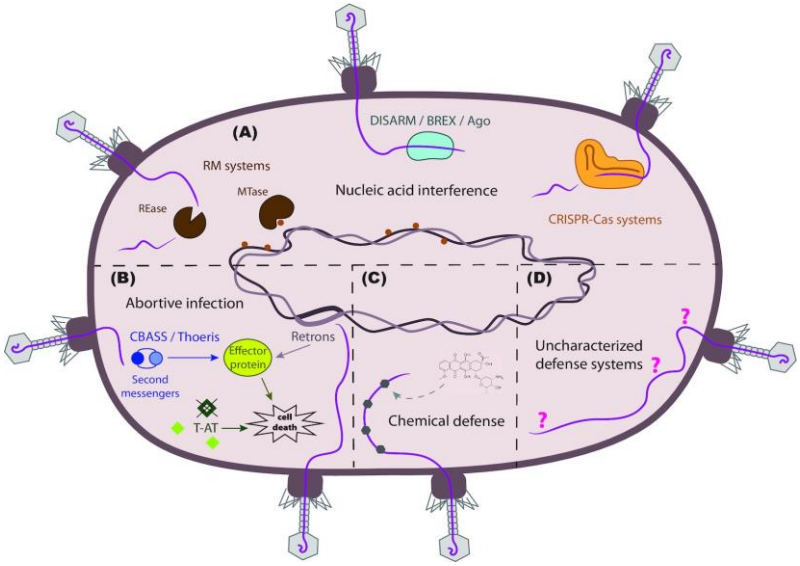

Figure 1.10. Post-adsorptive resistance mechanisms. (A) Mechanisms of nucleic acid interference (RM, CRISPR-Cas, DISAR, BREX, and Ago systems). **(B)** Mechanisms of Abortive infection (CBASS, Thoeris, Retrons, Toxin-Antitoxin (T-AT) systems) **(C)** Mechanisms of chemical defense mediated by secondary metabolites as daunorubicin which intercalate with phage DNA, preventing replication **(D)** Uncharacterized defenses system which mechanisms are still to be discovered. Source: [90].

1.2.4 Evolutionary and ecological consequences

Given the selective pressure of phages, bacterial surface receptors, and defense systems evolve and are gained and lost at a rapid pace [29,112]. Concomitantly, phage RBPs and counter-defenses also evolve fast to prevent phage extinction [22,113]. The diversification of bacteria and phages is thus the most obvious consequence of phage host-range. **Coevolution** is the process by which two or more entities undergo reciprocal, adaptive genetic changes [114]. There are two main hypotheses of phage-host coevolution: **arms race dynamics** (ARD) and **fluctuating selection dynamics** (FSD)[115].

In ARD, bacteria and phages continuously evolve toward new resistance and infectivity, without losing their ability to resist or infect the ancestral phage or bacteria, respectively. ARD is driven by **directional selection** and is associated with the **gene-for-gene** genetic model of coevolution [116]. In the latter, the acquisition

of phage genes or mutations increases infectivity, favoring host range expansion [117](Figure 1.11).

On the other hand, in FSD, phages evolve to overcome bacterial resistance, but at the cost of losing their capacity to infect ancestral hosts. Concomitantly, bacterial resistance to the contemporary phage compromises previous resistance to the ancestral virus. FSD is driven by **frequency-dependent negative selection**, where the frequency of alleles and their associated fitness change over time [118]. FSD leads to high specialization of phage-bacteria interactions and is associated with the **matching allele** genetic model [116] (Figure 1.11).

The interactions between phages and hosts can be visualized as a **network**, in which phages and bacteria are nodes connected by edges. It can also be viewed as a **matrix** with hosts on one axis and viruses on the other (Figure 1.11). Two main properties of networks are nestedness and modularity, which can be applied to phage-bacteria interactions [119]. Nestedness refers to a hierarchy of infection patterns (i.e., from specialist to generalist) and resistance. Nestedness is a result of the ARD or gene-for-gene model, in which the host-range of evolved phages is a subset of the host-range of ancestral phages. Similarly, resistance of evolved bacteria is a subset of the resistance of ancestral bacteria. In contrast, modularity is observed when phages and bacteria preferentially cross-infect within modules and is more a product of the matching allele model. One phage infecting one bacterium (one-to-one) in the most extreme case of modularity [117](Figure 1.11).

Nestedness is related to the evolution of **broad host-range phages** and modularity to specialized, **narrow-range phages.** The conceptual difference between the terms "broad" and "narrow", despite their widespread usage, is unclear and depends on the context in which they are used. The majority of described phages can infect a few strains of a particular bacterial species. Phages that can infect multiple species or genera have been described, but to a lesser extent [87]. It is likely that the host-range of phages in nature goes from extremely narrow to broad [120]. Metagenomics studies and alternative isolation techniques, which have shown that broad host-range phages are more common than previously thought, support this [41,121].

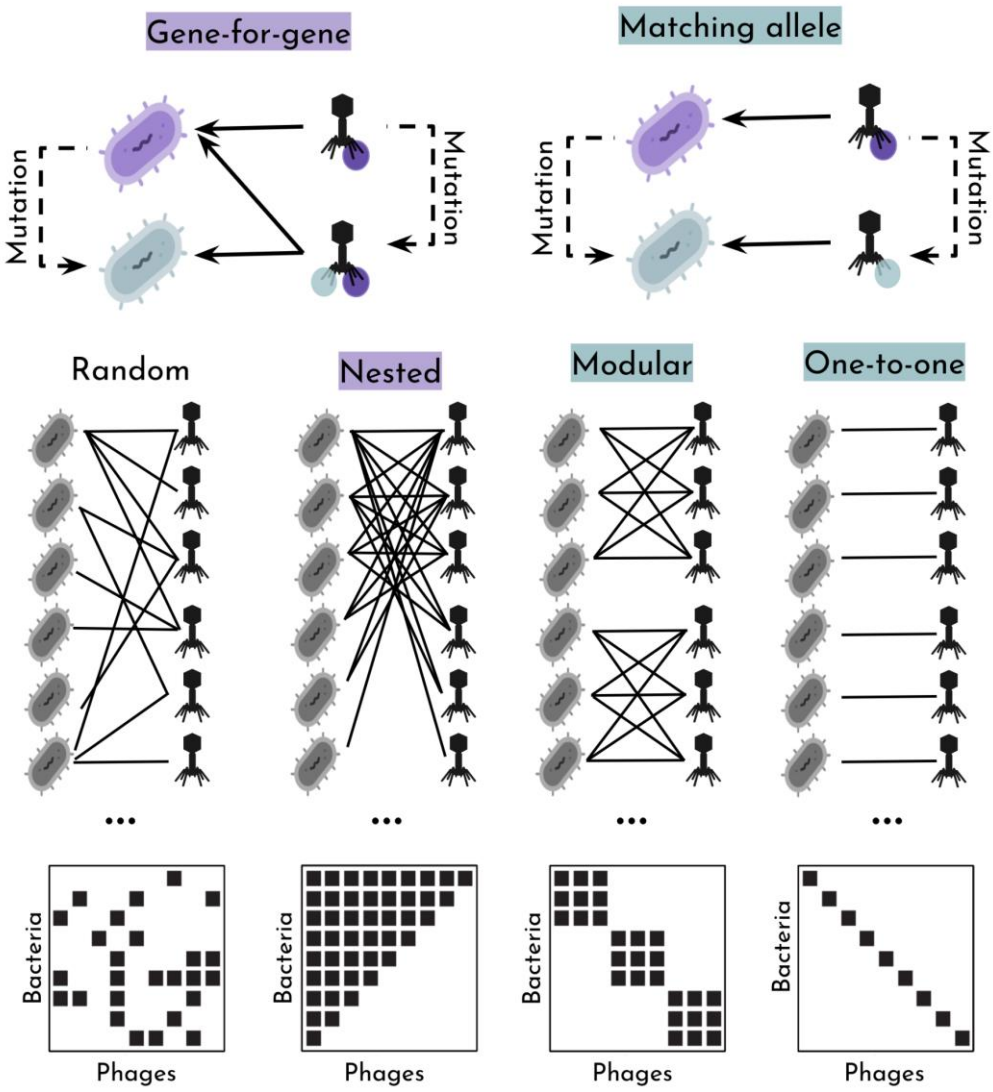

Figure 1.11. Models of phage-bacteria genetic coevolution and how these are translated into bacteria-phage infection networks and matrices. Edges and colored boxes in the networks or matrixes, respectively, represent infectivity. Source: adapted from [117].

Finally, the evolution of **cross-resistance** will also have an impact on the modularity or nestedness of bacteria-phage interaction networks [122]. Cross-resistance is the process in which resistance to one phage causes resistance to another phage. This is primarily due to shared receptors between phages. On the

other hand, resistance to one phage could also cause **cross-sensitization** to others, resulting in a trade-off [122]. Our limited knowledge of phage receptors and how different phages share strategies limits our ability to predict patterns of cross-resistance [86]. Given the difficulties of studying phage-bacteria interactions in nature, controlled laboratory experiments provide a starting point to better understand the arms race between bacteria and their parasites.

1.2.5 Studying phage-bacteria interactions

Because there are so many different genetic and phenotypic determinants at play, studying phage host-range is extremely challenging. This is worsened by the ongoing co-evolution of viruses and hosts, which makes host-range a transient trait. Besides, the experimental techniques used do not capture all the possible interactions between viruses and their hosts. Spot, plaquing, or liquid assays are the most widely used methods (**§1.2.1**). However, they are time-consuming and low-throughput. As they involve individual testing of each possible phage and bacteria pair, the actual host range of a phage is likely to extend to bacteria that could not be tested. In addition, they rely on both bacteria and phages being amended to cultivation conditions [87]. Biases also result from the preferred isolation of virulent phages, capable of completing their lytic cycle [123]. Alternative approaches, such as viral tagging [124], phageFISH [125], single-cell genomics [126], microfluidic PCR [127], proximity-ligation sequencing [128] or AdsorpSeq [129], have recently emerged. Despite these methods overcoming some of the limitations of the conventional assays, they are currently labor-intensive, expensive, and difficult to interpret [123,130,131]. Therefore, traditional methods continue to be the preferred option for studying phage-bacteria interactions in the case of culturable bacteria.

In addition to experimental techniques, *in silico* approaches can also be used to study phage-bacteria interactions [123,130]. These have become more important in recent years due to the development of high-throughput sequencing and bioinformatics tools. To predict phage host-range, 15 distinct methods or strategies have been developed, and 10 of these have been published in the previous two years [123]. First, bacterial and viral abundance profiles from metagenomics can link phages to hosts [132]. Second, phage-bacteria

coevolution leaves traces in the genetic makeup of bacteria, such as CRISPR spacers or prophages. Therefore, it is possible to predict a past infection event using sequence identity with spacers or prophages. Unfortunately, these methods cannot be applied to bacteria that are not lysogenic or do not have a CRISPR-Cas system. Also, spacers and prophages can accumulate mutations or may be lost during bacterial evolution [133,134]. Another option is to use co-evolutionary signals that are independent of sequence identity. For example, phages and their hosts frequently share oligonucleotide frequencies or patterns of codon usage [123,130]. In this regard, machine-learning (ML) methods use a combination of phage and host genomes properties, called 'features'. Features can range from nucleotide signatures to physico-chemical composition of amino acids or to the presence of protein domains [135,136]. Although the biological interpretation of ML methods is not straightforward, they are very promising. For example, a ML approach based on RBPs of phages has been recently developed at the species level for *Staphylococcus aureus*, *Klebsiella pneumoniae*, *A. baumannii*, *Pseudomonas aeruginosa*, *Escherichia coli*, *Salmonella enterica* and *Clostridium difficile* [137].

The lack of comprehensive phage and host databases is one of the main limitations of all *in silico* tools. For example, CRISPR spacers had a 15% accuracy and 13% sensitivity for predicting the host range of phages at the genus level [130]. Precision increased to 69% after mining of 11 million CRISPR spacers from more than 350,000 bacterial genomes [138]. Overall, available tools can predict phages' host-range with moderately good accuracy at the family and genus levels, but they perform worse at lower taxonomic ranks [139–141]. This is primarily caused by the scarcity of data on phage-bacteria interactions and the absence of systematic whole-genome datasets of tested pairs [142]. Therefore, experimental datasets are needed to train *in silico* predictive tools and improve our understanding of the biological patterns that underlie their interactions [136]. The utility of combining both experimental and genomic analyses has recently been addressed in *Bacteroides thetaiotaomicron,* a prevalent human gut bacterium, *Vibrio sp.*, an important marine bacterium, and *E. coli* K-12*, a* preferred model bacterium in biology [54,82,113,143]. However, curated datasets of phage-bacteria interactions are still necessary for most hosts.

1.3. Interplay between capsules and phages: *Klebsiella* as a model

1.3.1 The bacterial capsule

Despite the great diversity of phage receptors, the extracellular capsule is the outermost layer of many bacteria encountered by phages. Capsules are polysaccharides consisting of repeating units of 4-6 sugars (such as galactose, glucose, rhamnose, fructose, and mannose) with a uronic acid (such as galacturonic acid, glucuronic acid, or pyruvic acid). **Capsular polysaccharides (CPS)** are synthesized, transported, and firmly anchored in the cell wall envelope, creating a thick extracellular layer [144]. There are only a few examples of proteinaceous capsules such as the one produced by *Bacillus anthracis* composed of poly-γ-d-glutamate [145]. All bacterial phyla with sufficient genomic representation (> 20 complete genomes) have a capsule system. In particular, more than 70% of Gammaproteobacteria genomes encode a capsule [146]. This class includes ecologically, clinically, and scientifically important families of bacteria such as *Enterobacteriaceae*, *Vibrionaceae*, and *Pseudomonadaceae*.

The ubiquity of capsules is the result of their major role in bacterial ecology, evolution, and pathogenesis. First, capsules are associated with increased environmental breadth, as they allow colonization of new niches [147]. Second, capsules affect the rate of DNA exchange, with encapsulated bacteria being faster-evolving organisms [148,149]. Finally, capsules are important virulence factors, as they protect bacteria from desiccation, host immune responses [150], or chemical aggressions, including antibiotics [151]. For example, in a mouse model of pneumonia, acapsular *K. pneumoniae* cannot colonize and cause disease [152]. Consequently, CPSs have been a vaccination target, because anti-capsular antibodies can protect against disease [153]. In contrast, hyperencapsulated bacteria are associated with bloodstream infections and have increased dissemination and mortality rates [154]. Interestingly, the capsule can be detrimental in rich environments, for invasion of epithelial cells, or in urinary tract infections (UTIs). This could lead to the prevalence of mixed encapsulated and acapsular bacteria, both *in vitro* and *in vivo* [154,155].

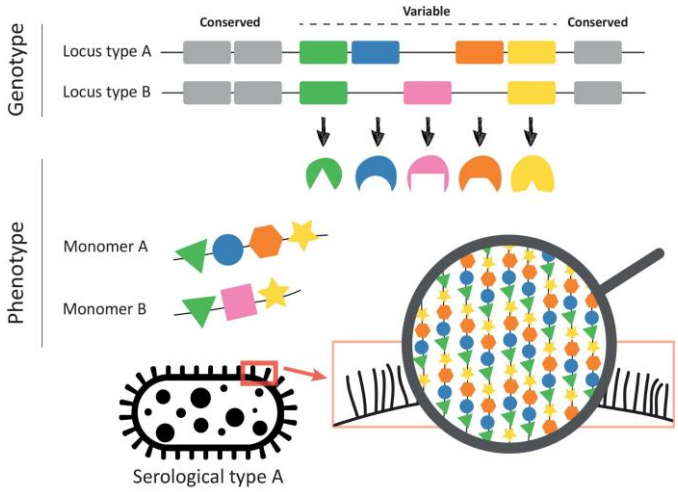

Figure 1.12. Genotypic basis of CPS phenotypic diversity. Schematic representation of the CPS genotype-to-phenotype path. The genetic sequence of the capsular locus is highly predictive of capsular serotype. However, it should be noted that (i) additional genes outside the capsular operon may affect the capsule composition, (ii) point mutations may change enzyme specificity, then producing a new serotype, (iii) information on cross-reactivity of CLTs is not available, (iv) many strains are not typeable as some CLTs are unknown. Source: [29].

Most CPS are synthesized through the Group I or Wzx/Wzy pathway. This pathway is located on a large chromosomal operon (10-30 kb) composed of variable genes flanked by a conserved region (Figure 1.12). The transport, assembly, and secretion machinery, as well as the enzymes necessary for the synthesis and modification of the specific sugars, are encoded in the central variable region, thus defining the capsular serological type (**serotype**)[29]. In *Citrobacter* spp. and *Salmonella* spp., capsular serotypes are referred to as Vi antigens but in most well-studied genera, such as *Escherichia*, *Klebsiella*, or *Acinetobacter*, they are known as **K-antigens**. Different K-antigens were defined using antisera. With the increased availability of genome sequences, *in silico* serotyping has now become the norm. For this purpose, several tools have been developed which assign a **capsular locus type** (CLT) based on the analysis of several genes or the entire capsular operon (see a review of the tools here: [29]) (Figure 1.12). For example, capsular type 1 can be classified as K1 (K-antigen 1) by serology or as KL1 by *in silico* serotyping (K locus type 1).

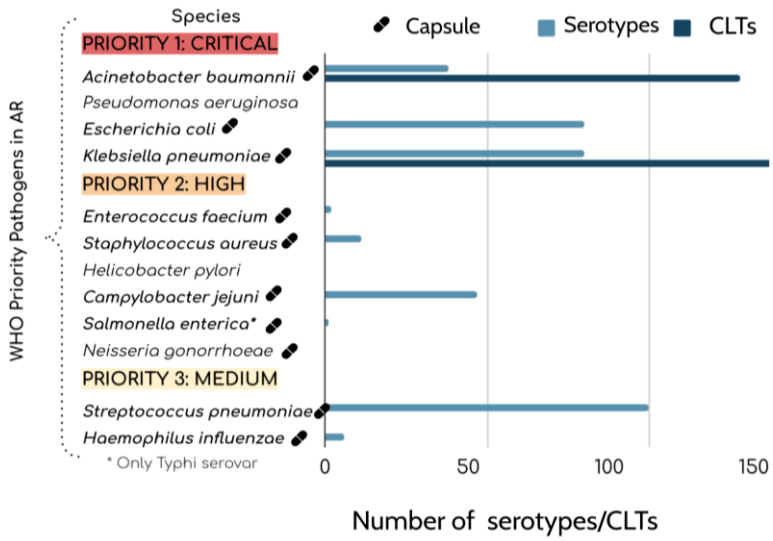

Figure 1.13. CPS diversity in multidrug-resistant pathogens. Diversity of capsular serotypes (K-antigens) and capsular locus types (CLTs) for priority pathogens declared by the World Health Organization (WHO) based on their antibiotic resistance. All species contain a CPS except *P. aeruginosa* which produces EPS alginate which is not anchored to the cell wall. Source: This work. Data extracted from [29].

The phenotypic and genetic diversity of the CPS depends on the bacterial species. For example, *S. aureus* has 11 serotypes, whereas for *E. coli* and *K. pneumoniae* about 80 K-antigens have been described (Figure 1.13). This difference is even greater if we look at CLTs. The most extreme case is *K. pneumoniae* in which more than 180 CLTs are known (Figure 1.13), with this number continuously growing. In fact, the diversity of the CPS exceeds ten-fold the diversity found at the O-antigen locus of the LPS, the other main surface antigen in *Klebsiella* [156]. This suggests that the CPS is undergoing diversifying selection in *Klebsiella* spp. as well as in other bacterial species [157,158]. This hotspot of genetic diversity is the result of *de novo* mutations, transposable element-mediated rearrangements, HGT, or homologous recombination [157,159]. These processes can modify the capsular structure or lead to capsule inactivation. Importantly, homologous recombination between conserved flanking regions could result in a '**capsule swap**' in which the capsular locus is replaced by another [149,158]. In addition to polysaccharides' evolvability, these structures also exhibit

increased robustness (that is, functional tolerance to mutations), which, in turn, also promotes diversity [160]. For example, polysaccharides are less constrained than protein antigens, which could cause fitness costs for bacteria with a single amino acid change [29]. Selective agents that further promote CPS diversification include host immune systems, cell-cell interactions, and phages that use the capsule as a receptor [29].

1.3.2 Phage capsular depolymerases

Given the ubiquity of bacterial capsules, phages that infect encapsulated bacteria have evolved strategies to overcome this barrier. Specifically, they encode enzymatic domains in their RBPs called **depolymerases (Dpos)** that cleave the O-glycosidic bonds of polysaccharides into soluble oligosaccharides. The existence of phage plaques with surrounding haloes led to the first description of Dpos in 1956 [161]. While the size of the plaque remains constant, the size of the haloes increases with incubation time (**Figure 1.14**). This is attributed to the diffusion of phage progeny, which could depolymerize but not lyse non-actively growing bacteria.

Dpos can be categorized as hydrolases or lyases depending on whether a water molecule is released during the substrate cleavage. Examples of hydrolases include sialidases, levanases, glycanases, and dextranases. Pectin-pectate or alginate lyases are other examples of Dpos [162]. These enzymes can have activity against the CPS, EPS, or O-polysaccharides of the LPS, and are typically found in the TSPs, being a structural element of the virion (**§1.2.2**). Alternatively, a few examples of Dpos have been detected on phages' baseplates or neck appendages [162]. In addition, some Dpos are thought to be soluble and released upon bacterial lysis rather than incorporated into the phage particle [163,164]. Here, we will focus on TSPs with Dpos that recognize, bind to, and enzymatically degrade the CPS.

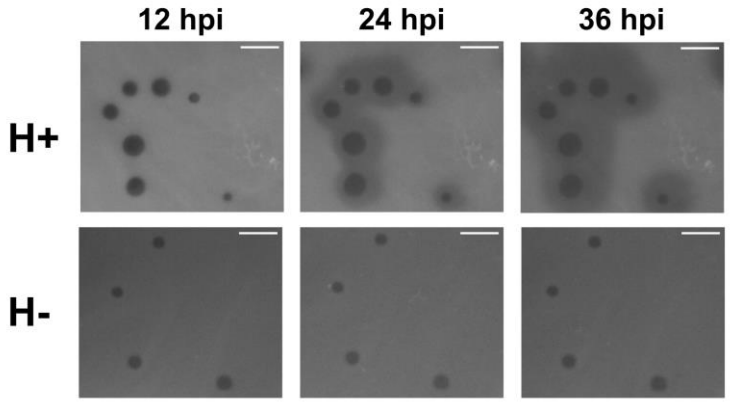

Figure 1.14. Morphology of phage plaques surrounded by haloes (H+) or not (H-). hpi: hours post-infection. Scale bar: 5mm. Source: This work.

TSPs with Dpos can have any similarity in their nucleotide or amino acid sequences, but their structure is highly conserved (Figure 1.7). Typically, this structure has three domains:

(i) A structural N-terminal domain that anchors the protein to the phage baseplate or tail or to other TSPs. This domain consists of a few hundred (150-300) amino acids but can be as short as 7-29 *aa*. Shorter domains are named conserved peptides and are thought to be sufficient to mediate attachment [165] (Figure 1.15).

(ii) A Dpo central domain responsible for host recognition and enzymatic activity. This domain is characterized by the presence of parallel β-sheets which make Dpos highly stable. These enzymes have a high degree of substrate specificity which is conferred by their primary sequence. If a particular cleavage site is present in different serotypes, a Dpo can be active against several **cross-reactive capsular types** [59,166].

(iii) A C-terminal domain that can contain a chaperone domain, involved in folding, or lectin-like domains that could also be involved in host recognition [83,163].

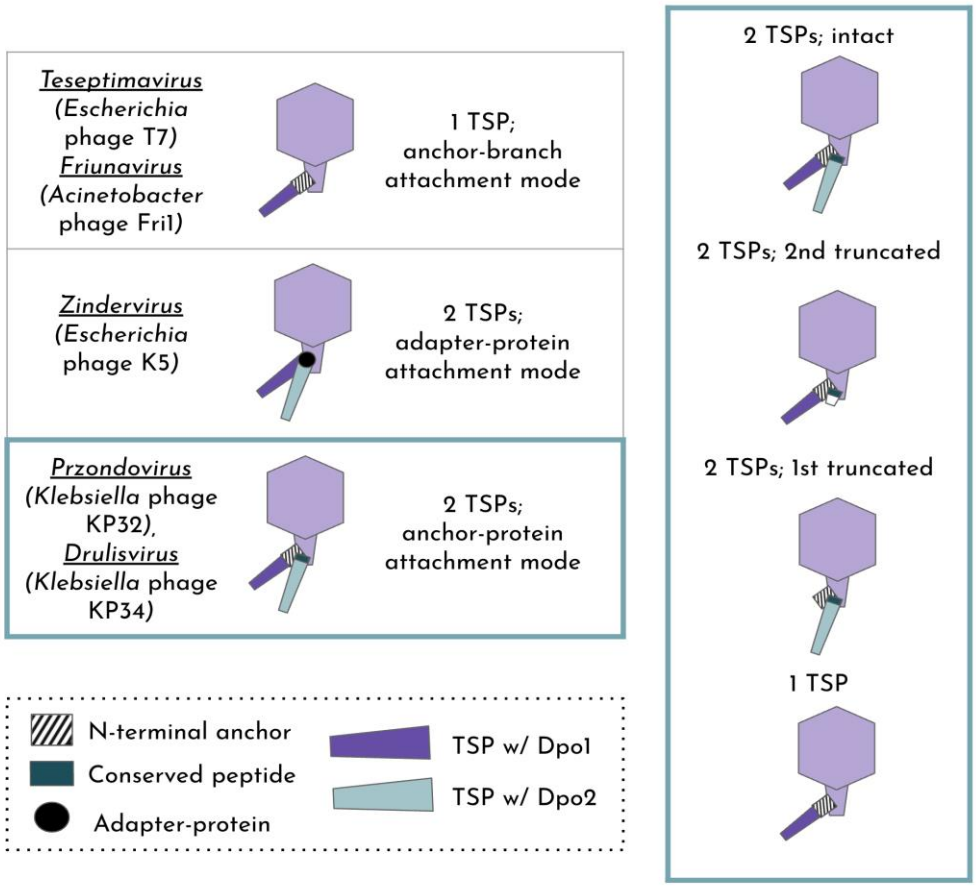

Figure 1.15. Phages with Dpo activity. Architecture of podovirus TSPs with Dpo activity. Left: examples of phage genera and representative phage species with their model of TSP attachment. Right: Types of conformations by phages with 2 TSPs**.** Only one copy of TSP is shown for simplification. Source: This work. Data extracted from [163,164].

In addition to the accumulation of mutations by vertical inheritance, the modularity of TSPs allows for rapid evolution by HGT [167]. While the anchor domain is conserved among phylogenetically related phages, the domains involved in recognition and degradation are highly variable. For example, Oliveira *et al*. characterized 12 podoviruses from distant geographic areas that differed only in the C-terminal pectate lyase domain of the TSP, having different specificities [168]. Most of the capsular depolymerizing phages (**Dpo-phages**) described are in fact podoviruses, which may require more degradative activity than phages with longer tails [79,168,169].

Dpo-phages are not “jack of all trades”; rather, there is a strong physical limitation on the depolymerases they can carry and/or express simultaneously. Podovirus have up to 2 TSPs with Dpos, allowing them to infect 1-2 capsular types or a few more in case of cross-reaction. The primary TSP is attached to the tail by the N-terminal anchor and provides additional branching sites for secondary TSPs [170]. The latter are linked via shorter peptides (**Figure 1.15**). In some cases, the central and C-terminal regions appear truncated even though the anchor/peptide is present (**Figure 1.15**). This might allow phages to rapidly acquire novel specificities by HGT, which may be advantageous in environments with fluctuating serotypes [163]. On the other side, some siphoviruses and myoviruses are also equipped with Dpos. They might have two branched-attached TSPs, similar to podoviruses [163](**Figure 1.15**). More striking is the case of jumbo myoviruses, such as *Klebsiella* phage ΦK64-1 or *Klebsiella* phage ϕKp24, which have an hyperbranched adsorption apparatus with 11 and 14 TSPs with Dpos, respectively. These allow these phages to infect 9 capsular serotypes and probably additional ones that have not been identified yet [170,171].

Although Dpo swapping allows phages to change capsular specificities [172], currently, it is not possible to determine the specificity of an enzyme from its primary sequence. Instead, the typical procedure involves extensive host-range testing along with the expression and purification of the candidate enzyme. This process requires the previous bioinformatic detection of the TSPs, which is also challenging [149,173]. Taken together, comprehensive analyses of the TSPs' diversity and how this diversity shapes capsular specificity are still lacking.

1.3.3 Capsule-phage interactions

Dpos not only mediate phage specificity, but they can also be involved in reversible and/or irreversible adsorption. The role of the CPS and Dpos for the subsequent steps of infection has only been described for a few phage-host pairs [174,175]. Since capsule inactivation is a main cause of phage resistance, it is widely recognized that Dpo-phages require the CPS for infection [91,173,176]. Most likely, phage-capsule interactions do not represent an all-or-nothing process. As a simplification, we propose four possible types of phage-capsule interactions (**Figure 1.16**):

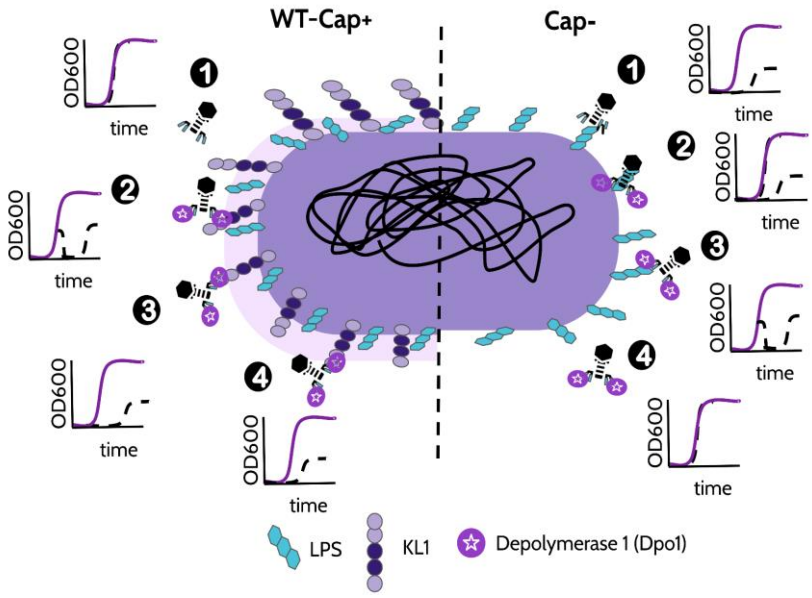

Figure 1.16. Schematic models of phage-capsule interactions and their dynamics. Phages without Dpo (1) or with Dpo (2, 3, and 4) are shown. WT-Cap+: wild-type encapsulated bacteria. Cap-: acapsular bacteria. Bacterial growth curves built from optical identity measurements (OD600) are shown to exemplify how we can study these interactions in the laboratory by the corresponding dynamics. Lines in purple indicate untreated bacteria and black, dashed lines, bacteria treated with the corresponding phage type. Source: This work.

(1) Non-Dpo-phages: If a phage lacks a Dpo enzyme, it could not penetrate the CPS, which hinders access to inner phage receptors. In this case, the capsule would prevent phage infection and removal of this layer might allow those phages to infect, expanding their host range [177,178].

(2) Dpo-phages: Degradation of CPS by Dpos provides phages with access to their inner receptors. In this case, the capsule does not prevent phage infection, but instead lengthens the time needed for the phage to adsorb. Then, the removal of this layer would be beneficial for phage dynamics [179].

(3) Dpo-phages and CPS as facultative receptor: The binding of CPS by Dpos helps the phage to initiate irreversible adsorption. Here, the capsule would increase the stability and positioning of the phage to find its secondary receptor and removal of this layer would reduce phage adsorption [175,180].

(4) Dpo-phages and CPS as obligate receptor: The binding of CPS by Dpo CPS is necessary to initiate irreversible adsorption. In this case, the capsule would be essential for phage infection, and removal of this layer would prevent phage adsorption and, thereby, infection [176,181].

Establishing the role of capsules and Dpos for infection is not trivial, as it can affect the co-evolution of bacteria and phages. Given the importance of the capsule in bacterial ecology, evolution, and pathogenesis (§1.3.1), phage resistance mediated by the CPS could have important fitness consequences [91,182]. For example, interactions of type (1) and (2) might promote hypercapsule production as a defense mechanism [94]. On the other hand, interactions (3) and (4) can select for down-regulation or loss of the bacterial capsule [173,183,184]. This can also result in ARD or FSD dynamics (§ 1.2.4), fostering the diversification or fluctuation of capsular and Dpos types, respectively. Not only that, but shifts in the encapsulation state of bacteria and the cargo of Dpos by phages are also expected due to their cost [149,155,179,185].

The evolutionary consequences of phage-capsule interactions are difficult to predict. First, the interactions could be influenced by environmental conditions or by chance [155,169,186,187]. Second, they could be affected by the **genetic background** of both bacteria and phages [188]. Most of our knowledge of phage-capsule interactions comes from individual phage-bacteria pairs [175,189], which precludes deciphering more general patterns. This is especially relevant considering the huge diversity of phages and bacteria on Earth, and their mechanisms of infection and defense, respectively. Therefore, the generality of the previous results to other bacteria and phage systems remains to be ascertained.

1.3.4 *Klebsiella pneumoniae*

K. pneumoniae is a particularly interesting model to study the interplay between phages and capsules:

(i) *K. pneumoniae* is a member of the clinically and scientifically relevant *Enterobacteriaceae* family. The model organism *E. coli,* which is closely related to *Shigella spp.* [190], and other important pathogens such as *Salmonella*, *Serratia*, *Enterobacter*, *Raoultella* and *Yersinia* are also members of this family. Other

species such as *K. quasipneumoniae and K. variicola* are also included in the denominated *K. pneumoniae* species complex group while *K. oxytoca*, *K. terrigena*, *or K. spallanzanii* are more distantly related [191].

(ii) *K. pneumoniae* is encapsulated via the ubiquitous Group I or Wzx/Wzy pathway and it is the species with the highest observed capsular diversity (§1.3.1). Sequences and atomic structures of 75 reference K-serotypes are available [192]. Additionally, a tool for genotypic typing of *Klebsiella* called Kleborate allows for the direct determination of CLTs from genomic assemblies [156,159]. This has opened new possibilities for addressing questions related to *Klebsiella* CPS diversity and the processes shaping it [149,193].

(iii) *K. pneumoniae* genomes are extremely diverse. *K. pneumoniae* typically encodes 5,000-5,500 genes and only 2,000 are core genes, i.e., shared by all members of the species. In fact, the proteome of *K. pneumoniae* exceeds 100,000 different protein-coding sequences [194]. Several schemes are used to distinguish between lineages [195]. The most common is the Multi-Locus Sequence Typing (MLST) in which allelic profiles or Sequence Types (STs) are defined based on the sequence of seven housekeeping genes [196]. As of July 2022, more than 6,000 different STs have been reported [197]. In addition to STs, clonal groups (CGs) are used to study *Klebsiella* population structure. Phylogenetic patristic distances (threshold = 0.04) were used to define 259 CGs, of which 28 are responsible for the majority of *K. pneumoniae* isolates reported [198]. Beyond the capsule, hotspots of *K. pneumoniae* genome diversity include plasmids [199], prophages [173], and intracellular defense systems [98], all of which contribute to phage infectivity [113,200].

(iv) *K. pneumoniae* is a main target for phage therapy because of its clinical significance and increasing burden of antimicrobial resistance. *K. pneumoniae* is widely distributed and can be isolated as a free-living organism in soil, water, food, plants, etc. Additionally, it can also be found in wild and domestic animals such as birds, insects, cows, and horses. Principally, *K. pneumoniae* is found in the microbiome of the mouth, skin, and gut of healthy individuals as commensal bacteria [201,202]. Otherwise, *K. pneumoniae* can cause opportunistic infections such as pneumonia, urinary tract infections, diarrhea, meningitis, or sepsis. Environmental and clinical strains of *K. pneumoniae* do not represent separate

subpopulations [194,203]. Instead, important clinical lineages of *K. pneumoniae* have been isolated from non-human sources [203]. As a result, it is possible to isolate phages from environmental samples against clinical strains, which are easier to obtain than free-living bacteria.

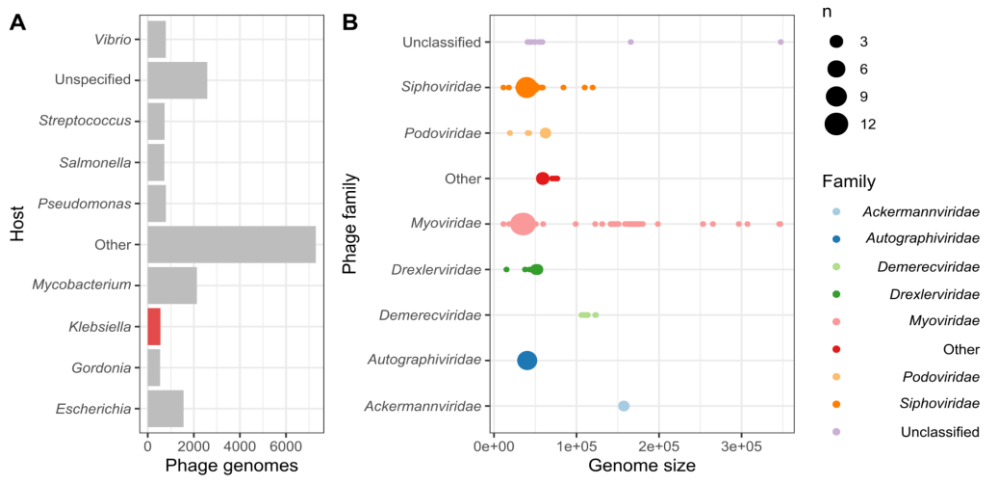

Figure 1.17. *Klebsiella* phages available. A. Number of phage genomes per bacterial host. 'Other' includes bacterial genera targeted by less than 3% of the phages. **B.** *Klebsiella* phages genome size grouped by family. The size of the points is proportional to the number of genomes. Source: This work. Data extracted from https://github.com/RyanCook94/inphared (July 2022).

1.3.5 *Klebsiella* phages

Phages infecting *Klebsiella* have been isolated from a variety of sources such as sewage, seawater, and animal feces [204]. As of July 2022, there are over 555 *Klebsiella* phage sequences available, up 8-fold from five years ago. This number is still lower than for other relevant genera such as *Mycobacterium*, *Escherichia*, *Pseudomonas*, or *Staphylococcus* (Figure 1.17.A). *Klebsiella* phages range in size from 11,444 to 347,546 (median: 44,923) and belong to seven families of the Caudovirales order. Specifically, the most common families are *Autographiviridae* (n=161), *Myoviridae* (n=175), and *Siphoviridae* (n=68) (Figure 1.17.B). Rarefaction analysis of phages showed that the number of novel species and genera continues to increase with the number of genomes analyzed [39]. This suggests that we have only scratched the surface of *Klebsiella* phage diversity and that new phages remain to be discovered.

The exponential growth of *Klebsiella* phage genomes is uncoupled from their biological characterization, as only 25% of phages have been published with information beyond their genome sequence (Beamud *et al.*, unpublished data). Thus, although this increase expands our knowledge of phage diversity, little is known about the general patterns of phage-bacteria interactions [82]. *Klebsiella* phages are equipped with Dpos and, as such, exhibit capsular specificity (§1.3.2). For instance, using a common K22 host, we isolated 6 closely related *Klebsiella* phages [59,205]. These phages had a narrow host-range when tested against the collection of 77 K-serotypes of *K. pneumoniae*. Specifically, all phages (VLC1-6) infected K22 and K37. The capsular loci encoding for K22/K37 differ in a point mutation, and their atomic structures are almost identical (Figure 1.18). In contrast, phage VLC6 encodes an additional Dpo (Dpo 3) that confers this phage an extended spectrum toward the cross-reacting K2 and K13 capsular types [166], referred then as K2/13. Phage VLC6 can also depolymerize but not lyse bacteria from K3. This ability could be conferred by both Dpo1 and Dpo3 given the overlap of the K3 antigen structure with K22/37 and K2/13 [59] (Figure 1.18).

To date, *Klebsiella* phages against 44 different capsular types have been characterized. Within these, 23 have been experimentally correlated with their corresponding Dpo (Table S1). Considering the large diversity of *Klebsiella* capsular surfaces, new Dpos remain to be found. Some of these may be active against capsules not yet typeable. Furthermore, most phages have been tested using just one strain of the corresponding capsular type [59,171,206]. This raises the question of how infection is correlated in other bacteria with the same capsule. On the other hand, a few *Klebsiella* phages without detectable Dpos have been reported [173,178,206,207]. This could be due to divergent Dpo sequences or alternative mechanisms of action. In any case, their host tropism and RBPs remain largely unexplored.

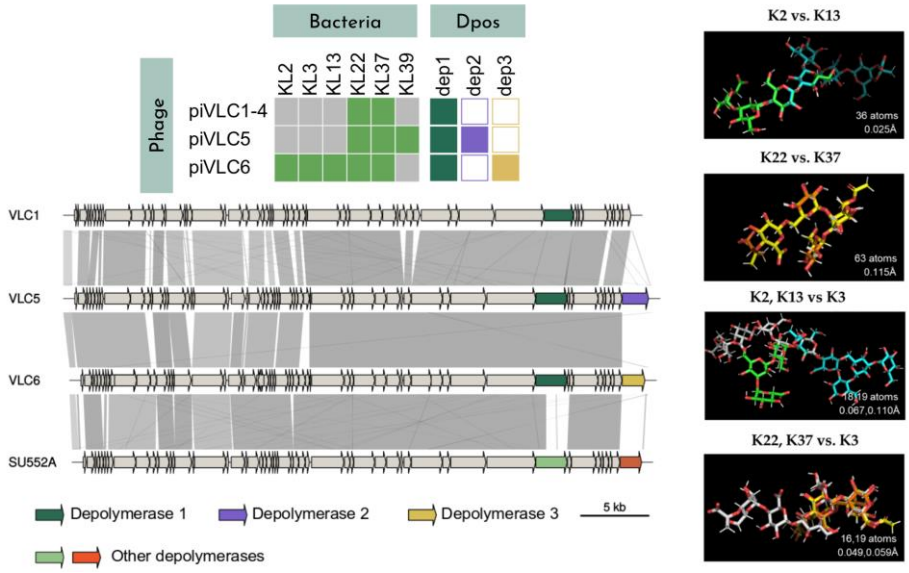

Figure 1.18. Host-tropism of VLC *Klebsiella* drulisviruses. Left: Infection profile and Dpo dosage of *Klebsiella* phages VLC1-4, VLC5, and VLC6. The genome comparisons of VLC and SU552A phages are shown. The VLC1 genome is representative of those of phages VLC2-4. SU552A was the closest genome to VLC phages from the NCBI. Source: This work. Right: Structural superposition of some *Klebsiella* K-antigens to show cross-reactivity. The number of atoms that align between the pairwise structures and the root-mean-square deviation (RMSD) values are indicated. Green—K2 (85 atoms), white—K3 (112 atoms), blue—K13 (112 atoms), yellow—K22 (98 atoms), and orange—K37 (93 atoms). Source: Modified from [59].

2. Objectives

This thesis aims at getting a deeper insight into the determinants of phage-bacteria interactions in *K. pneumoniae*, with emphasis on the interplay between capsules and depolymerases and the effect of both bacteria and phage genomic backgrounds. For this purpose, we combined phenotypic analysis with genomic data. The specific goals are the following:

1. **To build and characterize a representative collection of *K. pneumoniae* and *Klebsiella* phages to study host-range:**

    1.1. To select *K. pneumoniae* isolates representative of the capsular diversity of the species.
    1.2. To sequence and characterize previously isolated *Klebsiella* phages.
    1.3. To compare the diversity of *Klebsiella* strains and *Klebsiella* phages with bacteria and phages from the databases.

2. **To study the relative importance of different bacteria and phage genetic traits as determinants of phage host-range and infection outcome:**

    2.1. To perform all-vs-all analyses of interactions between bacteria and phages.
    2.2. To quantify the sensitivity of bacterial and phage traits shaping the infection matrix.
    2.3. To perform sequence-based host-range predictions.

3. **To determine the role of the capsule for successful phage infection and phage dynamics:**

    3.1. To isolate a collection of acapsular bacteria.
    3.2. To characterize the infection of acapsular bacteria by phages.
    3.3. To quantify the dynamics of bacteria and phages.

3. Material & methods

3.1 Bacteria and phages

3.1.1. Bacteria

Bacterial strains used in this study are detailed in **Table S2**. The Networked Laboratory for Surveillance of Antimicrobial Resistance (NLSAR) of Comunitat Valenciana [208] and InfectERA Consortium [209] provided access to bacterial strains and their genome sequences. Genome sequences were obtained by Illumina NextSeq sequencing. All strains were isolated from clinical specimens from 8 hospitals of the Comunidad Valenciana, Spain. Jin Town Wang provided the NTUH-K2044 strain, whose genome was publicly available [210].

3.1.2. Phages

Phages used in this work are detailed in **Table S3**. Phage plaques were isolated and provided by the Environmental and Biomedical Virology Lab (I2SysBio, UV-CSIC). Phage isolation was done as previously described [205] from soil and water samples near a sewage plant in Valencia, Spain (**Figure 3.1**). Clinical isolates of *K. pneumoniae* available at that time were provided as hosts (**Table S2**). Phages were stored at 4 °C in Luria Bertani (LB) broth or SM buffer (100 mM NaCl, 8 mM $MgSO_4 \cdot 7H_2O$, and 50 mM Tris-Cl pH 7.5) and routinely titered and amplified if necessary.

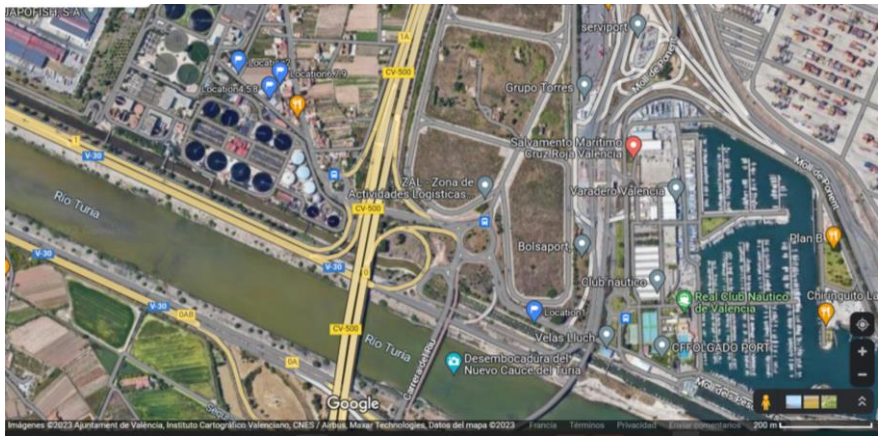

Figure 3.1. Location of environmental samples used for phage isolation. Sampling points are labeled with a blue flag.

3.2 Bacteria laboratory methods

3.2.1. Culture conditions

Bacteria were grown in LB agar plates or in LB broth. Culture conditions were 37 ºC, including shaking (150 rpm) for liquid cultures. Liquid cultures and plaques were incubated overnight for up to 48 h. A single colony was used to make 25% glycerol stocks that were stored at -80 ºC for further use. Unless otherwise specified, bacteria were grown directly from frozen stocks with a pipette tip in LB broth.

3.2.2 DNA extraction and *wzi* PCR

Single colonies used for bacterial stocks were resuspended in 100 µL of sterile water and DNA was extracted by boiling (99 °C, 10 min). Lysates were then centrifuged (18,000 x g, 10 min) and DNA was recovered from the supernatant and stored at −20 °C until use. PCR amplification of the *wzi* gene, a conserved but hypervariable gene of the capsular operon, was performed with modifications [211]. As new sequences have been deposited since then, new *wzi* primers were designed. The reference database of *Klebsiella* capsular types was downloaded from Kaptive [159], and the *wzi* gene was extracted using a semantic search. Gene sequences were aligned with **MAFFT** v7.271 and conserved regions were detected with **Aliview** [212]. Annealing characteristics of new primers wzi2_F (5'-ACAATGATRAAAATYGCGCG-3') and wzi2_R (5'-GAGAGCCACTGGTTCCAGAA-3') were verified with the **Benchling Primer Design** interface (https://benchling.com/). The reaction volume was 25 µL, containing 1x PCR buffer, 200 µM dNTP mixture, 0.5 µM of each primer, 0.5 U Taq DNA polymerase (Biotools), and 2 µL of DNA. Cycling conditions were 94 °C for 5 min followed by 30 cycles of 94 °C for 1 min, 53 °C 40 s, 72 °C 30 s, and 20 cycles of 94 °C 10 s, 55 °C 30 s, 68 °C 3 min, and a final extension step of 72 °C for 5 min. Amplification was confirmed by electrophoresis by loading 5 µL of the PCR reaction on a 1.6% agarose gel. Positive PCR products were purified using the NucleoFast 96 PCR clean-up kit (Macherey–Nagel).

3.2.3 Sanger *wzi* sequencing

Purified amplicons were sequenced using the BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems) using the same PCR primers. Two Sanger sequencing reactions were set up for each PCR reaction, one for sequencing the forward strand and the other for the reverse strand. Given the overlapping between the two resulting sequences, this allows increasing the confidence of each base called. Briefly, the reaction mix contained 1.6 µL of buffer, 0.4 µL of Big Dye Terminator, 1µL of primer at a final concentration of 0.625µM, and 1 µL of the purified *wzi* PCR product. The total volume of the reaction was 8 µL. The Sanger reaction was incubated at 95 ºC for 5 min, followed by 60 cycles of 95 ºC 10 s, 50 ºC for 5 s, and 60 ºC for 4 min. The samples were then sent to the Central Service for Experimental Research of the University of Valencia, which performed capillary electrophoresis using an ABI 3730xl DNA sequencer (Roche).

To obtain the consensus sequences from the electropherogram files, the make.consensus.seqs function of the R package **sangeranalyseR** [213] was used *(min.reads = 2, trim.cutoff = 1e-04, min.length = 20, secondary.peak.ratio = 0.33)*. If necessary, manual inspection of electropherograms was performed with the **Staden package** [214]. Sanger *wzi* sequences were aligned with the previous *wzi* database (**§3.2.2**) using **MAFFT** v7.271 *(--add)*. Then, a phylogeny was built using **IQ-TREE** v.1.6.5 with 1,000 ultra-fast bootstrap replicates. The capsular type of bacteria was assigned based on the closest relatives approach. This consisted in extracting the set of sequences that grouped with the *wzi* sequence of interest with bootstrap support higher than 70% [65].

3.2.4 Capsule extraction and quantification

The bacterial capsule was extracted as previously described [215] with modifications. 2 mL of overnight cultures were mixed with 100 µL of capsule extraction buffer (500 mM citric acid pH 2.0, 1% Zwittergent 3-10) and vortexed. The mix was heated at 56˚C for 20 min and cell debris was removed by centrifugation (18,000 x *g*, 5 min). To precipitate the capsule, the supernatant (400 µL) was mixed with absolute ethanol to a final concentration of 80% (v/v) and incubated for 30 min at 4 °C. The precipitates were collected by centrifugation

(18,000 x *g*, 20 min, 4 ˚C), air-dried, and resuspended in 200 µL of water. Before quantification, CPS was incubated for 2 h at 56 °C [155]. CPS was quantified by the uronic acid method [216]. Briefly, the capsule suspension was mixed with 1,200 µL of borax solution (12.5 mM disodium tetraborate in H_2SO_4), kept on ice for 10 min, followed by incubation at 95 °C for 10 min, and immediate cooling on ice for another 10 min. The absorbance of the sample at 520 nm was measured after the addition of 20 µL of 3-hydroxybiphenyl in 0.5% NaOH. Measures were done within 1 hour in quadruplicate in a Tecan Infinite 200 Pro and the median was obtained. A standard curve of glucuronic acid (5 µg/mL - 50 µg/mL) was used to calculate uronic acid concentrations (µg per 200 µL of capsule suspension) for each quantification batch. As a negative control, 200 µL of water instead of capsule suspension was quantified in parallel. If required, concentrations were normalized to the number of colony-forming units (CFUs).

3.2.5 Isolation of acapsular (Cap-) bacteria

Colonies of *K. pneumoniae* have a mucoid-opaque morphology and spontaneous acapsular (Cap-) mutants can be distinguished as translucent colonies or as clearer sectors within a colony. To isolate Cap- mutants, overnight cultures (c.a. 1.5e+9 CFU/mL) were diluted and plated in Tryptic Soy Agar (TSA). Since TSA plates are darker than LB plates, there is a greater contrast to help distinguish between different capsular phenotypes. Concretely, cultures were diluted 10,000-fold, and 2 µL of the suspension was pipetted into plates to obtain c.a. 300 individual colonies. This yielded a theoretical detection limit of Cap- of 5.00E+06 CFU/mL (CFU/mL=$\frac{1\ x\ 10{,}000}{0.002\ mL}$) assuming 1 Cap- at the 10^{-4} dilution. Bacteria were mixed with 100 µL of Phosphate Buffer Saline (PBS) and the mixture was evenly distributed with the help of glass beads. Plates were incubated for 24 hours and Cap- mutants were determined by the naked eye and light contrast. Cap- and total colonies on plates with ≥ 100 CFU were counted to obtain the number of Cap- per 100 CFU of WT bacteria.

Cap- mutants were streaked-purified in LB plates to confirm phenotype stability. Stable Cap- mutants were grown in 1 mL of LB to make 25% glycerol stocks and stored at -80 ºC. A maximum of three colonies per independent culture was selected to avoid a jackpot effect [217], i.e., that all Cap- mutants derive from a single early mutational event. In addition, capsule production was checked by the uronic method assay for a subset of Cap- (**§3.2.4**).

3.3 Genome analyses of bacteria

3.3.1. De novo assembly and typing

Bacteria included in this work were previously sequenced using Illumina NextSeq 500 with the Nextera XT library kit. FASTQ files were assembled *de novo* with **SPAdes v3.9.1** [218], in the case of the analysis of surveillance sequences, or with **Unicycler v.0.4.8** [219] for the case of sequences selected for the Kpn-collection. This was due to the higher number of sequences in the former and the computational requirements of each program. For strains of the Kpn-collection, raw reads were also filtered and trimmed prior to assembly with **PRINSEQ-lite v.0.20.4** [220]. Reads with a mean quality lower than 25, read length shorter than 60 nt, larger than 500 nt, or optical duplicates were removed. In addition, the last three positions of the 3' end and positions with a quality lower than 20 were trimmed. The resulting bacterial assemblies were annotated with **Prokka v1.14** [221].

Kleborate v.0.3.0 [222] was used to obtain the assembly quality (number of contigs and N50), species, sequence type (ST), capsular locus type (CLT), O antigen locus type (OLT), virulence score, antibiotic resistance score, number of resistant antibiotic classes, and number of resistant genes of bacteria. For a detailed description of Kleborate workflow and results, see https://github.com/katholt/Kleborate/wiki.

3.3.2. Phylogenetic analyses

For building phylogenetic trees, the core genome alignment of the bacterial isolates was obtained first. To do so, **Snippy v.4.6.0** (https://github.com/tseemann/snippy) *(--ctgs option)* was used with the NTUH-

K2044 chromosome as reference (accession: NC_012731). Positions with gaps in at least 10% of sequences were excluded from the alignment with **trimAl** [223]. **IQ-TREE2** [224] was used to infer a maximum-likelihood phylogeny [225] under the best fitting model [226]. If sequences from the NCBI database were included in the phylogenies, these were previously obtained with **ncbi-acc-download** (https://github.com/kblin/ncbi-acc-download) using their respective accession numbers.

3.3.3 Identification of phage receptors

Verified phage receptors were obtained from the **phageReceptor** database [80], and sequences were downloaded from the NCBI using **efetch** [227]. Additionally, gene IDs were saved and gene synonyms were retrieved from **UniprotKB** [228]. This resulted in a list of bacterial proteins and gene names that served for phage receptor searching (**Table S4**). Two strategies were used: (i) based on sequence identity with verified receptors and (ii) based on functional annotation. For the first strategy, receptor protein sequences were concatenated and this file, as well as the bacterial proteomes (*.faa files from Prokka, **§3.3.1**), were used as input for **PIRATE** [229] (75-80 amino acid identity % thresholds). This program identifies the presence or absence of orthologous gene families, which included the verified phage receptors. For the second strategy, a semantic search was performed using gene IDs and synonyms. These were queried against each ortholog's consensus product assigned by **PIRATE**. Orthologs that had "outer membrane" in their assigned product were also included. Both strategies were combined resulting in a presence/absence matrix of putative proteins involved in phage recognition.

3.3.4 Identification of phage defense systems

For the detection of homoimmunity, i.e. resident prophages that may contain genetic elements identical or nearly identical to those of the incoming phage that may give rise to an immunity phenotype, prophages were first identified with **PhageBoost** [230]. Then, identity of phages to prophages was determined with **blastn** *(e-value 1e-5, min identity 80%, and alignment length >= 99 nt)*. For the detection of CRISPR immunity, **CRISPRCasTyper** [231] was used. CRISPR

spacers were extracted and compared against phages with **blastn-short** *(e-value 1e-5, min identity, and coverage 90%)* [173]. The remaining defense systems were detected using hidden Markov models (HMMs) profiles. HMMs are probabilistic models that consider the conservation of each specific amino acid in a multiple sequence alignment (MSA) and allow the prediction of protein structures and functions when the primary sequence is not conserved. Two strategies were used:

(i) A custom pipeline. HMMs of RM and CBASS systems were obtained from https://gitlab.pasteur.fr/erocha/RMS_scripts and [232], respectively. Pfam identifiers of the rest of defense systems were obtained from [232]. HMMs profiles were extracted from the **Pfam-A v.33** database using **hmmfetch** [233]. Bacterial hits to the defense systems HMMs were obtained with **hmmsearch** [233]. Hits were filtered using an e-value and coverage thresholds of 0.001 and 0.8, respectively, with the **RMS_script** (https://gitlab.pasteur.fr/erocha/RMS_scripts/). The output was a set of bacterial proteins with significant hits to any of the domains found in the verified defense systems.

(ii) **Defense-finder** [98]. This pipeline uses MacSyFinder [234] after HMM searching to retain only those hits that satisfy the genetic architecture of each defense system as co-localization of several proteins is often required to have a functional system. The output was a list of confident systems in a given bacterial assembly. Unless otherwise stated, results from this strategy are shown.

3.4 Phage laboratory methods

3.4.1 Phage titration

Phage titers were obtained by a slightly modified tube-free soft agar overlay assay [235]. Briefly, 0.5 mL of an overnight culture of the isolation host was poured over an LB plate dish (bottom agar). 2 mL of 0.3% LB agar were then added to the plate (top agar) that were swirled to mix the bacteria and to spread it across the plate. Plates were allowed to dry for 5 min and then 2 µL of 10-fold serial dilutions of phages (10^{-1}:10^{-8}) were spotted onto the bacterial lawn and allowed to dry for 10 min. Plates were incubated at 37 °C and the next day phage plaques were counted from the lowest dilution with individual plaques. The counts were then transformed to Plaque-Forming Units per mL (PFU/mL).

3.4.2 Phage amplification and purification

Phages were amplified in 5 mL of exponential cultures of the isolation host (c.a. 10^7 CFU/mL) in LB broth for 3 h. After centrifugation (13,000 × g, 5 min), supernatants were passed through a 0.22 µm syringe filter to remove bacteria. In case a drop in phage titer was observed, supernatants were passed through 0.45 µm, or filtration was replaced by a double centrifugation step. Two methods were tested for obtaining high-titer purified lysates for sequencing, Polyethylene glycol (PEG) precipitation and ultrafiltration. For PEG, 4 mL of phage lysate was mixed by inversion with pre-chilled PEG-NaCl 3X in 15 mL centrifuge tubes, resulting in a final concentration of PEG 8000 10% and 1M NaCl. The mix was incubated at 4 ºC overnight to allow the virions to precipitate. Phages were pelleted by centrifugation for 1 h at 15,500 x g at 4 ºC. For ultrafiltration, lysates were loaded into the upper reservoir of an Amicon filter device (100 KDa). Manufacturer instructions were followed including two washing steps with 4 mL of SM buffer. PEG-precipitated and ultrafiltrate-purified phages were recovered in 200 µL of SM buffer and titrated to ensure successful phage DNA extraction (>= 10^9 PFU/mL) [236]. Gel electrophoresis (in 0.8% agarose) and Qubit (high-sensitivity dsDNA Kit) were used to quantify free DNA from the host.

3.4.3 Phage morphology

For transmission electron microscopy (TEM), a drop of each high-titer purified phage was sent to the Centro de Investigación Príncipe Felipe microscopy service. Phages were deposited onto a carbon-coated Formvar supported by a 300 mesh copper grid and air-dried for 30 min. Viruses were negatively stained with 2% phosphotungstic acid and visualized under Jeol JEM-1010.

To record phage plaque morphology, molten streaking from singles [235] was used. This method was chosen as it allows to directly obtain individual plaques regardless of the titer of the input phage stock. Briefly, 0.5 mL of an overnight culture of the isolation host was poured over an LB plate dish (bottom agar), and then 2 mL of 0.3% LB agar was then added to the plate (top agar) that was swirled to mix the bacteria and to spread it across the plate. Without letting it dry, a loop was used to streak viruses from the purified lysate and to make a repeating Z-stroke (**Figure 3.2**). Plates were allowed to dry for 20 min and then incubated at

37 °C. After 24 h, plaque morphology was recorded for well-separated plaques, and photographs were taken on a Vilver E-box. Phage plaques and haloes were measured using **ImageJ** [237] for at least 6 plaques per phage.

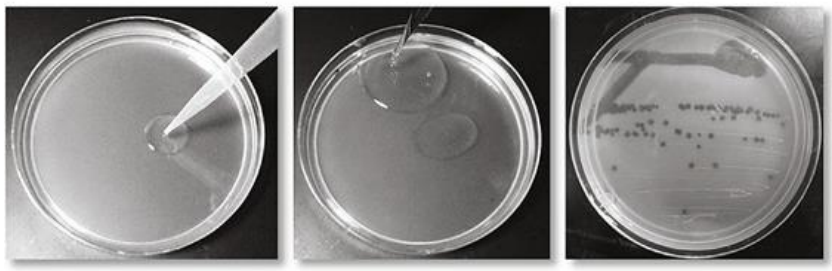

Figure 3.2. Procedure to obtain single plaques and record morphology. From the left to the right: pipetting of overnight culture, addition of top-agar with a serological pipette, and streaking of phages. Source : [235]

3.4.4 Phage sequencing

Purified phages were ten-fold diluted in Turbo DNAse Buffer 1X (final volume of 100 uL) and 2 µL DNAse Turbo, benzonase, and micrococcal nuclease were added. The digestion was incubated for 1 h at 37 °C and 1 µL of DNase Turbo was added for another hour. Degradation of non-encapsulated DNA from the host was checked by Qubit and gel electrophoresis. Then, to inactivate DNases, 15 mM EDTA was added and the mix was incubated at 75 °C for 10 min.

To digest phage capsids, 10 µL of SDS 10% and 5 µL of proteinase K (20mg/mL) were added and incubated for 45' at 55 °C. The DNA Clean & Concentrator 5-Kit (Zymo) was used to extract and purify the phage DNA which was eluted in 12 µL of ultra-pure water. Purified DNA was quantified with Qubit (high-sensitivity dsDNA Qubit Kit) and normalized to 0.2 ng/µL. Extractions with the lowest performance (low ratio phage DNA: host DNA) were repeated. This resulted in several phages with DNA extraction and sequencing replicates (referred to as technical duplicates) that served for genome assembly optimization afterward. Sequencing libraries were prepared using the Illumina Nextera XT DNA kit and paired-end reads (2 x 250 bp) were generated in the Illumina MiSeq platform (MiSeq Reagent Kit v2). Because transposases cannot be integrated at the ends, the use of a tagmentation kit to fragment the DNA and ligate the adapters

will prevent the identification of phage genome ends from sequencing coverage [238]. However, due to DNA requirements, alternate kits that rely on the random fragmentation of phage DNA would have required more time (eg. TruSeq DNA Nano needs 100 ng as input).

3.5 Genome analyses of phages

3.5.1 *De novo* assembly

First, the quality of the resulting sequencing reads was evaluated using **FastQC** [239] and summarized with **multiQC** [240]. Taxonomical classification of reads to detect host contamination was performed with **Kraken2** *(database: minikraken2_v1_8GB)*[241] and summarized with **Pavian** [242]. Raw reads were filtered and trimmed with **PRINSEQ-lite** [220] *(-min_len 20 -min_qual_mean 20 -ns_max_n 30 -trim_qual_right 20 -derep 14)*. If possible, assembly was done with **Unicycler v0.4.7** [219,220] as more consistent results were observed between technical duplicates. If the complete genome was not obtained, **SPAdes v.3.14** [243] or **MIRA4** (https://sourceforge.net/projects/mira-assembler/) were employed. For SPAdes, several kmer sizes were used for assembly, and the maximum size *(-k 127)* was selected as the optimum. For MIRA4, a previous subsampling of fastq reads was done using **seqtk sample** (https://github.com/lh3/seqtk) as an excess of coverage was detrimental to the assembly.

Most phage genomes are packaged as linear molecules but due to terminal repeats, the assembly process results in an apparently circular contig. Even artifactual, circularity generally indicates a complete phage genome [244]. The circularity of contigs was directly assessed from **Unicycler** or with the **apc script** (https://github.com/tseemann/apc/blob/master/apc.pl). Phage contigs were extracted based on several criteria: circularity, length (>1,000 nt), depth coverage (>100x), and corroboration of non-host origin after **blastn** with a custom bacteria database (Garcia-Lopez R. *et al.*, unpublished).

3.5.2 Clustering & classification

The clustering of phage genomes was done in two steps. First, **FastANI** [245] *(--fraglen=50)* was used to obtain the average nucleotide identity (ANI) between all phages. Pairs of phages with reciprocal ANI values of >=99% from >=99% of fragments were considered duplicates. After deduplication, intergenomic similarity values (IGS), more precise than ANI, were obtained with **VIRIDIC** [52]. IGS values were transformed into a matrix of genetic distances (100-IGS) that were used to build a neighbor-joining tree with the **ape package** [246].

To define similarity groups and species clusters, the ANI.dendrogram function of the **bactaxR package** [247] was used with IGS thresholds of 95% and 45%, respectively. Phage genomes available at NCBI and their metadata [39] (PhageDB) were obtained with **inphared** [39] (10/02/2021, n=14,037) and compared against isolated phages with **FastANI** *(--fraglen=500)*. PhageDB sequences with at least 70% ANI and 30% alignment coverage were kept for computing IGS values and built clusters with our phages as above. To each phage, a family, subfamily, and genera were assigned based on the available taxonomic classification of the phages from the PhageDB included in the same cluster. Phages were named with the first letter of the viral family, the number of the phage group, and the letter for identification within the group.

Finally, the diversity of the phages included in the Kpn-Phage collection was compared with the diversity of *Klebsiella* viruses described. From the PhageDB metadata, all the phages whose host was '*Klebsiella*' were extracted. The clustering of the Kpn-phages within the *Klebsiella* phages was investigated with **ViPTreeGen** [50,248] which uses a proteomic tree approach [47] capable of discerning between higher taxonomic ranks [50]. The resulting tree was visualized with **iTOL** [249].

3.5.3 Functional annotation

For annotating phage genomes, two strategies were used:

(i) Phage genomes were annotated with **Prokka v1.14** [221] with a custom database constructed from all phage proteins (July, 2020). The best **Viral Orthologous Group (pVOG)** of each protein was retrieved after a profile HMM

search with **hmmscan** *(--domlout option)* [233]. Results were parsed with **rhmmer** (https://github.com/arendsee/rhmmer) and filtered using a sequence e-value of 10^{-5} and profile coverage of 60%. Functional categories of pVOGs were retrieved [69]. Additionally, ORFs were searched against phage proteins with **blastp** (e-value cutoff of 10^{-5}). A consensus annotation was obtained excluding ambiguous and hypothetical proteins, if possible, which was used to define custom functional categories (1. Head, 2. Neck, 3. Portal, 4. Tail, 5. Coat, 6. DNA metabolism & Recombination, 7. Transcription, 8. Translation, 9. Other enzymatic, 10. Packaging, 11. Lysis, 12. Lysogeny, 13. Immunity, 14.Virulence, 15. Unknown). The obtained results were compared with those from pVOGs to exclude inconsistencies.

(ii) Phage genomes were annotated using **prokka** but with **PHROGS**, prokaryotic virus proteins clustered using remote homology [250]. This was done as suggested (http://millardlab.org/2021/11/21/phage-annotation-with-phrogs/). Unless otherwise stated, results from this strategy are shown to allow standardization.

For visualization of phage genomes and their functional annotation, phages were arbitrarily permuted at the small/large terminase subunit as is the convention when the genome termini are not known [244]. To do so, a custom script based on **tblastn** searches and **seqkit restart** [251] was used. Then, **blastn** comparisons between pairs of phage genomes were performed *(-evalue 1e-5)*. Finally, **genoplotR** [252] was used to draw comparisons of phage genomes as previously (https://github.com/BBeamud/phage-genomics/tree/master/plots).

3.5.4 Lifestyle prediction

Phage lifestyle was predicted with 3 different softwares: **PhageAI** [253], **PHACTS** [254], and **BACHPLIP** [255]. Lysogeny marker genes (integrases, excisionases, repressors, Cro/C1 domains, and *ParA*/B genes) [256,257] were also searched from custom functional annotation (strategy i) and from **PHROGS** (strategy ii) using semantic searching. Finally, **ABRicate** (https://github.com/tseemann/abricate) was used to screen for antimicrobial resistance (MEGARes database [258], December 2020) or virulence genes (VFDB [259], December 2020) in the phage genomes *(--minid 70 --mincov 60)*.

3.5.5 Identification of phage RBPs

A. With depolymerase activity

A sequence database of *in silico* or experimentally validated RBPs with Dpos was built from the literature (**Table S5**). All phage proteins (> 100 amino acids) were compared with this database using **blastp** *(-evalue 1e-5)*. Candidate proteins were verified to have both the N-terminal anchor and the central/C-terminal enzymatic domain using structural alignments. For the N-terminal domain, **hmmsearch** *(e-value: 0.01 and coverage 0.01)* was used against a database of HMMs N-terminal profiles obtained from the literature [163,260]. For the C-terminal enzymatic domain, **hhblits** *(database PDB70)* [261], **InterProScan5** [262], and **Phyre2** [263] were used. Proteins that structurally aligned with profiles that contained the terms 'depolymerase', 'pectin', 'pectate', 'sialidase', 'levanase', 'xylosidase', 'rhamnosidase', 'dextranase', 'alginate', 'hyaluronidase', 'hydrolase', 'lyase', 'lipase' and 'spike AND beta-helix|hydrolase' were kept for further analyses [162]. Additionally, **hhblits** were executed for every phage ORF. No additional Dpos apart from those obtained with the **blastp** search were found [162]. Proteins with likely enzymatic activity were confirmed to have the characteristic **β**-helical domain using **gget** (https://github.com/pachterlab/gget) **AlphaFold2** [264]. As such, they were referred to as tail spike proteins (TSPs).

B. Without depolymerase activity

For detecting RBPs independently of Dpo domains, the **PhageRBPdectection** tool [260] was used. Structures of putative RBPs were obtained with **gget** (https://github.com/pachterlab/gget) **AlphaFold2** [264] and queried against the AlphaFoldDB and PDB protein structures with **FoldSeek** [265]. Template modelling scores (TM-scores), were obtained to determine the similarity between candidate RBPs and the rest of protein structures [265].

3.5.6 Classification of RBPs systems

To determine how TSPs were incorporated into phage particles, a similar nomenclature to that of Latka *et al* [163] was followed. When an **N-terminal domain** was identified in the RBP, this was assigned as the '**primary RBP**'. When a **conserved peptide** was found, the RBP was assigned as the '**secondary RBP**'.

Both corresponded to RBPs 'anchor-attached' (§1.3.2). If the structural domains were missing or could not be identified, the RBP was classified as 'unknown' because it might represent a putative soluble protein. If a structural domain was present but the Dpo was absent, the RBP was classified as 'truncated'. If the enzymatic domain present was different from a typical Dpo, it was also specified.

Structural and enzymatic domains previously detected were used for this classification. Additionally, anchor domains and conserved peptides were identified based on MSAs of RBPs of related phages (RPBs-MSAs). RBPs-MSAs were built as follows: Each RBP was compared against all phage proteins of the PhageDB and proteins of our phages with **blastp** *(-evalue 1e-5 -max_hsps 1 -length 100)*. Hits considered as intra-genus were used to build a sequence alignment with **MAFFT** *(--adjustdirection)*. An intra-genus hit was considered when the corresponding phages scored at least 70% ANI in 40% of their genome with **FastANI**.

3.5.7 Identification and clustering of receptor binding domains

To focus on regions driving host specificity, anchor domains of Dpos were removed. The conserved region of RBPs-MSAs was detected using **Gblocks** [266] *(--b3 4)* (**Figure 3.3**). The number of non-conserved positions (b3 parameter) was chosen based on the number from 3 to 8 that selected a single block that spanned less than 50% of the protein length in more RBPs. If a block could not be assigned or the remaining non-conserved positions were less than 20 *aa*, the whole protein was kept. For 5 cases, the anchor domain was delimited from visual inspection of the alignment. After delimiting the anchor domain, the remaining length of the protein was extracted with **seqtk subseq** (https://github.com/lh3/seqtk), which we referred to as **receptor binding domain** (RBD).

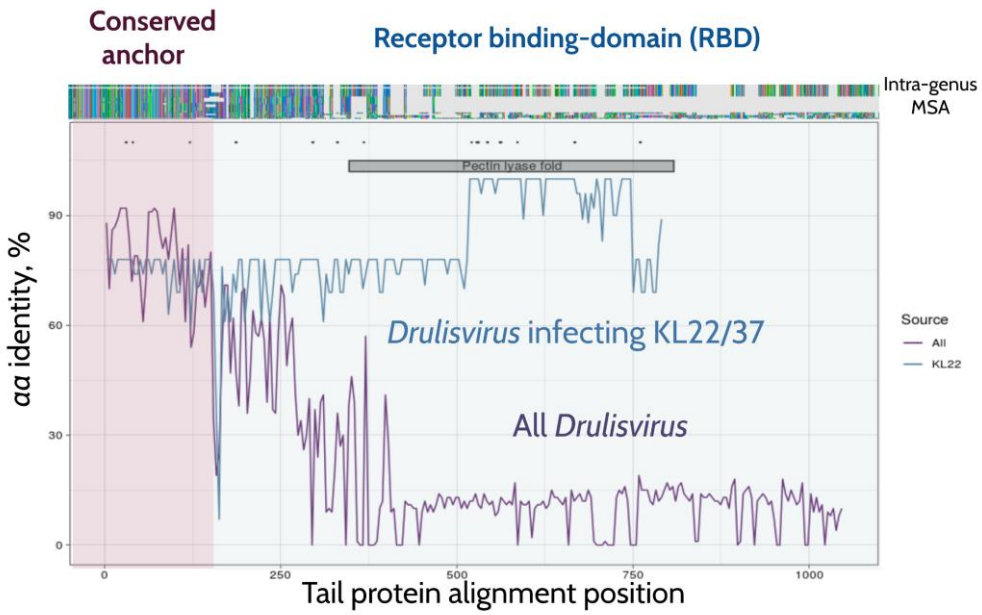

Figure 3.3. Example of an RBP-MSA and the delimitation of the anchor and receptor binding domain (RBD). The average amino acid identity was determined using the **alistat library** (http://caps.ncbs.res.in/iws/alistat_ali.html) after fragmenting the RBP-MSA into sliding windows of size 5 with a step of 4 *aa*. Black dots denote positions where a B-sheet was predicted with at least 80% probability using the **PredictHEC** function [267] of the **DECIPHER** R library. The gray bar indicates where **InterProScan5** or **HMMER** predicted a depolymerase domain. All: The alignment included all phages with the presence of the RBP within the genus. KL22: The alignment only included phages that infected the CLT KL22. Source: This work.

To determine clusters of RBDs several strategies were compared: (i) **mmseqs2** *(easy-cluster, --cov-mode 1)* [268] and (ii) all versus all **blastp** comparisons *(-evalue 1e-5 -max_hsps 1)* following by medoid clustering with the **bactaxR** package [247]. For the two strategies, several identities and coverage thresholds (40-80%) were tested. Clustering after blastp comparisons showed increased sensitivity (**§3.7.1**) and was the strategy finally used.

3.6 Phage-bacteria methods

3.6.1 Spotting assay

Spots assays consist in adding a phage drop to a bacterial lawn and observing areas of bacterial clearance after incubation (Figure 3.4). To do so, high-titer phage lysates were diluted ten-fold in a 96-deep-well plate (>10^7 PFU/mL) using a chessboard pattern. Phages were not adjusted to the same starting concentration to maximize the number of interactions observed for phages that reached higher titers and to allow for higher throughput [54,112]. Therefore, a 2 µL drop of each phage was added with a multichannel pipette to bacterial lawns in 0.3 % top-agar LB media (§3.4.1) and incubated overnight. A negative control with LB was included per plate. Three independent temporal blocks were performed with the order of the phages in the plate randomized. A spot was considered positive when at least two replicates showed a positive interaction (clear spot, turbid spot, or plaques).

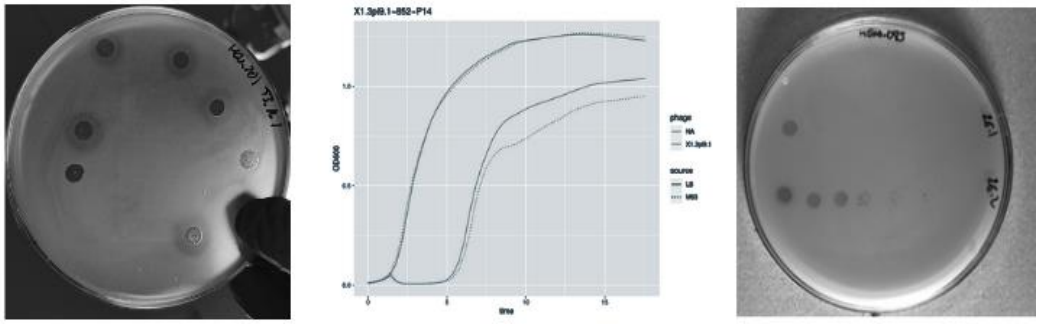

Figure 3.4. Examples of host-range assays. From left to right: 1. Spot-test assay. A phage-bacterium interaction was recorded as positive when in the area of the phage drop bacteria showed impaired growth. A negative control with SM buffer was always included. **2. Bactericidal assay**. A phage-bacterium interaction was recorded as positive when the bacterial growth curve in the presence of phage showed delayed or impaired growth compared to the bacteria-only control. **3. Dot-plaque assay**. A phage-bacterium interaction was recorded as positive when individual phage plaques were observed at greater dilutions of the phage spot. Source: This work.

3.6.2 Bactericidal assay

The ability of phages to lyse bacteria was tested by the planktonic killing assay. Briefly, bacteria were grown to the exponential phase, and cultures were adjusted to the same starting concentration after measuring the OD_{600}. Adjusted cultures of bacteria and phages were used to inoculate 200 µL of LB in a 96-well plate, resulting in 10^7 CFU/mL and 10^8 PFU/mL, respectively. Plates were then incubated at 37 °C with shaking (60 rpm) in a Tecan Infinite 200 Pro and OD_{600} was recorded every 20' for at least 20h. A positive bactericidal effect was assigned when a delay or change in bacterial growth was observed between the control and phage-treated bacteria in at least two replicates (Figure 3.4). When high bacterial concentrations are used coupled with inefficient or low phage titers, the bacterial growth is mostly unrestricted and cannot be differentiated from the growth kinetics of the control, resulting in a false negative. To avoid this, the inoculum size and phage titer used was carefully selected based on previous work (Figure 3.5.). This choice was also verified by testing several concentrations of phages with their corresponding isolation hosts.

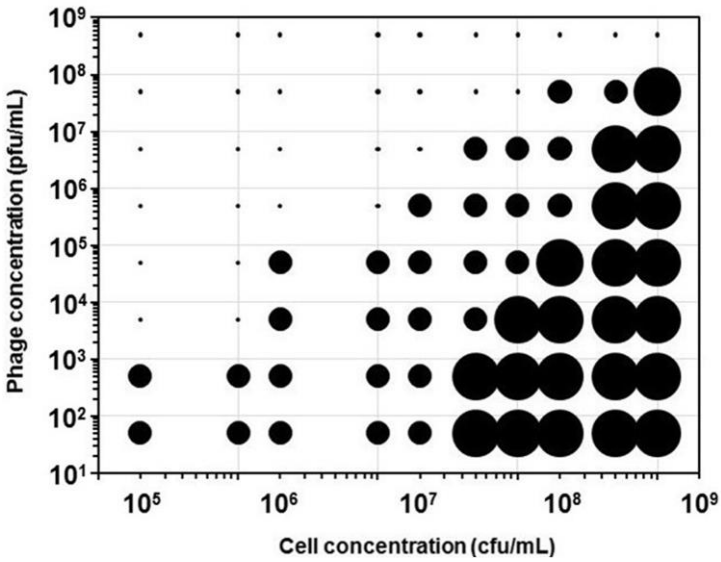

Figure 3.5. Detection of lysis using different concentrations of phage T4 and *E. coli*. Large circles indicate the absence of lysis, therefore false negatives of the assay. Intermediate circles indicate the existence of delayed lysis and small dots indicate complete lysis right from the start of the experiment. Source: [269].

3.6.3 Dot-plaque assay

The ability of phages to form plaques in bacteria was determined using the modified tube-free soft agar overlay assay as explained earlier (§3.4.1). Concretely, serial dilutions of phages from 10^8 PFU/mL to 10^4 PFU/mL were spotted in bacterial lawns. The next day, phages able to form individual plaques were recorded as positive cases for this assay. If a spot was observed at lower dilutions but individual plaques did not form, this assay was recorded as spotting positive but plaquing negative.

3.6.4 Progeny and adsorption assays

Phage progeny and adsorption assays were performed in a similar fashion, measuring the amount of free phage after co-incubation with bacteria. At least 3 individual colonies were picked and grown in LB broth until they reached 10^7 CFU/mL. Bacteria were then infected with ~10^5 PFU/mL and incubated for 3 hours at 37 °C with orbital shaking (250 rpm). Then, plates were centrifuged (4000 x *g*, 10 min) and phages were titrated as usual (§3.4.1). Free phage was calculated using the following expression: $100 \times \ 1 - \frac{A}{Ao}$, being A and Ao the phage PFU/mL with or without bacteria, respectively. A positive progeny production was considered if free-phage exceeded 10-fold over free-cell control. For combinations where phage was not totally adsorbed (free-phage > 10 PFU/mL), the incubation time was shortened to avoid interference with newly produced particles. Free phage was measured after co-incubation with its corresponding original host at 5, 10, 15, 20, 30, and 40 min to determine the best time to measure adsorption. 30 min was selected as the optimum as allowed to observe the maximum level of adsorption for all phages. The same procedure was used to determine irreversible adsorption but, after co-incubation of phage and bacteria, the mixture was vigorously vortexed and centrifuged (13,000 x *g*, 5 min). Positive adsorption was considered when the mean free-phage titer dropped > 80% compared to host-free controls.

3.6.5 Depolymerization assay

First, the presence of plaque haloes with increasing incubation times (12, 24, and 36 h) was recorded as indicative of phage depolymerization activity (**Figure 1.14**). Second, a new assay to quantify phage capsular-degrading activity in liquid was developed. As positive controls during the setup of this assay, several phage-bacteria combinations in which capsule reduction was checked by microscopy were used [59]. If phages have depolymerization activity, they will break the capsule into monomers that will be released into the medium. Briefly, aliquots of 2 mL of overnight bacterial cultures (c.a. 1E+09 CFU/mL) were treated with 10 µL of high-titer phage or 10 µL of SM buffer. Stationary cultures were used to ensure that CPS is not restored after cell division. Given that phages could not replicate in these conditions, high-titer lysates (≥ 1E+11 PFU/mL) were employed to ensure a ratio of phages to bacteria (Multiplicity of Infection, MOI) close to or greater than 1. The mix was incubated at 37 ºC with moderate shaking (150 rpm) to ensure phage-bacteria interactions. Next day, cultures were centrifuged (18,000 x *g*, 5 min) and washed twice with PBS. This step was necessary in order to remove CPS remnants in the supernatant and ensure the quantification of the remaining CPS anchored to the cells. Then, the capsule was extracted and quantified as described earlier by the uronic method (**§3.2.4**). If phages had non-degrading activity, the colorimetric assay resulted in a pink color, similar to the control of SM-buffer-treated bacteria. On the contrary, when phages had activity against the capsule, the pink color diminished or did not develop. Therefore, phage depolymerization activity was calculated using the following expression: $100 \times 1 - \frac{C}{Co}$, being C and C_o the amount of bacterial glucuronic acid (µg/200µL) with or without phage, respectively. This procedure was repeated at least twice for each phage-bacteria combination tested.

3.7 Additional bioinformatic and statistical approaches

3.7.1 True positive rate (TPR)

The TPR (i.e. statistical sensitivity) of each genetic trait to predict positive infections within infection matrices was calculated. In the case of bacteria, for each factor (host trait) and factor level (allele), a reference strain was selected and its pattern of infection determined. This infection profile was compared with the pattern of infection of other bacteria with the same allele, resulting in a contingency table of positive or negative infection given the outcome in the reference strain (Figure 3.6). From this table, the TPR was estimated as follows: $TPR = \frac{P(S|SR)}{P(S|R) + P(R|SR)}$, being S|SR sensitivity to phage given sensitivity in the reference bacteria and R|SR resistance to phage given sensitivity in the reference bacteria. The selection of the reference bacteria was repeated 100 times to avoid biases. To determine the null TPR distribution for each trait, the reference strain was randomly selected from the whole dataset. When the host trait represented binary data (eg. presence/absence of a receptor), the TPR was estimated only from the presence level. This process was repeated 3 times to estimate the standard error.

For the TPR of phage traits, each phage-CLT combination was considered instead of individual phage-bacteria pairs. We assumed that a phage was able to infect a CLT if at least one strain of the corresponding CLT was infected by the spotting assay or was referenced in the literature. Therefore, for each phage trait, a reference phage was selected and the positive/negative infection pattern of each CLT was compared with the infection of other phages from the same factor level (Figure 3.6).

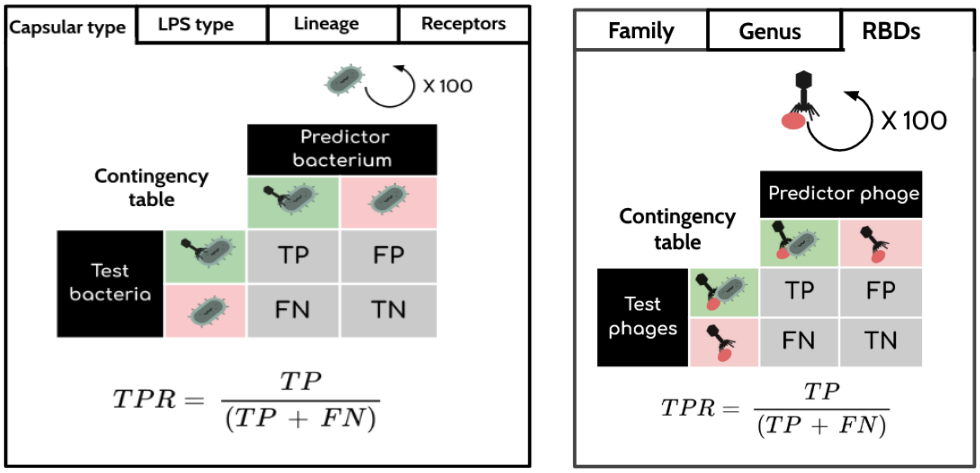

Figure 3.6. Procedure to evaluate the effect of bacteria (left) or phage (right) traits shaping the infection matrix of bacteria and capsular types, respectively. TPR: True Positive Rate. TP: True Positive. FP: False positive. FN: False negative. TN: True negative. Source: This work.

3.7.2. Cross-validation of RBDs with prophages

Representative sequences of RBD clusters were used to perform a search against the complete RefSeq protein database of bacteria with **blastp** *(-evalue 1e-5 -max_hsps 1 -qcovs 0.4)* to obtain putative prophage regions. Protein accessions of hits were used to retrieve the bacterial assembly accession and the corresponding taxonomy with **efetch**. Bacterial assemblies were downloaded from the NCBI ftp site and subjected to **Kleborate** [159] to determine the CLT of each assembly. Bacterial assemblies with 'None' or 'Low' confidence were removed from the analyses. Then, the CLT assigned to each RBD from the cluster analysis was compared to the CLT of the bacteria with the residing prophage (Figure 3.7). To test for statistical significance, a random CLT was assigned to each bacterial assembly with a probability equal to the frequency of each CLT in the *K. pneumoniae* RefSeq database (accessed: 27/04/2021). Then, Fisher's exact test was used to check the null hypothesis that RBDs do not have predictive power. To do so, the proportion of correct predictions in randomized versus actual data was performed.

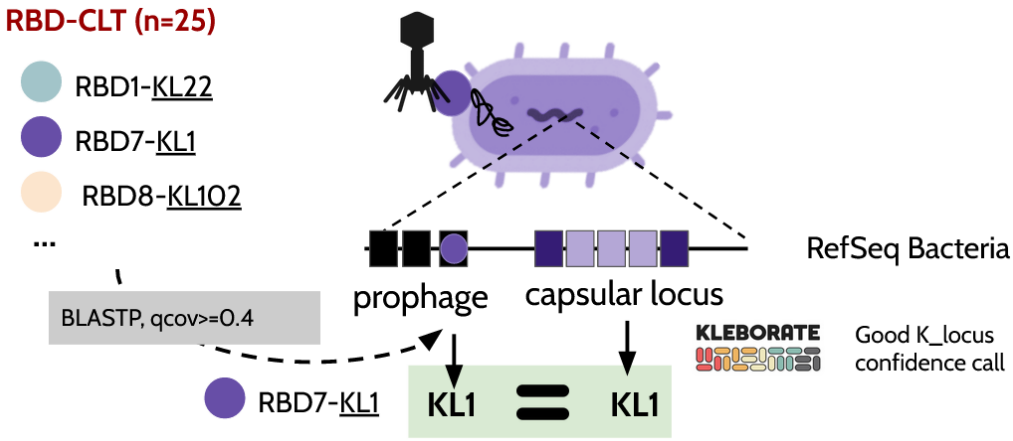

Figure 3.7. Scheme of the procedure to evaluate the capsular specificity of RBDs with prophage sequences. A representative sequence of candidate RBDs were queried against the Bacteria RefSeq searching for identity to residing prophages. Prophages represent past infectious events and should have recognized the bacterial receptors, which might include the capsular type of the host. If identity to an RBD was found in a prophage, a prediction was made about the capsular tropism of the temperate phage, here KL1. This prediction was compared to the capsular type of the lysogenized host, obtained with Kleborate. Source: This work.

3.7.3. Mechanisms of avirulent infections

Phages were compared by **blastn** with residing prophages *(e-value 1e-5, min identity 80%, and alignment length >= 99 nt)* to determine possible superinfection exclusion. **CRISPRCasTyper** [231] was used to identify CRISPR arrays which were compared against phages with **blastn-short** *(coverage and identity thresholds of 90% and e-value 1e-5)* [173]. The presence of each defense system (DF) given by **defense-finder** (§3.3.4) was compared with the outcome of phage virulence (V) or avirulence (A) in that particular strain. Given this, the **differential probability of resistance (dPR)** for each phage-defense system was calculated as follows: $dPR = P(DF|A) - P(DF|V)$, with values close to 1 indicating an overrepresentation of the DF in resistant strains (phage avirulence). On the contrary, negative values indicate that the DF was found overrepresented in susceptible strains (phage virulence) and they probably do not contribute to the observed resistant phenotype.

3.7.4 Analysis of growth curves

Bacterial growth curve measurement based on OD_{600} is one of the most commonly used methods in microbiology [270]. Comparison of bacterial growth in the presence or absence of phage can provide information not only about the host range of the virus (§3.6.2) but also about other properties such as phage virulence and resistance emergence [271,272]. Bacterial lysis curves are characterized by a first phase of absence or interruption of bacterial growth, after which resistance can emerge, resulting in regrowth of bacteria (Figure 3.4). In order to map the time of phage lysis and bacterial regrowth, we obtained the growth difference (ΔOD_{600}) between the phage-treated bacteria and the untreated control. ΔOD_{600} reflects phage-bacteria dynamics with values close to 0 indicating that, in the presence of phage, bacteria are growing similarly to the control, and negative values indicate a growth impairment due to the virus.

First, to determine the **time of phage lysis (t_{lys})**, ΔOD_{600} values of consecutive time points were adjusted to a linear model using the R lm() function and the coefficient extracted. T_{lys} was defined as the time point preceding the most negative slope in ΔOD_{600} values. Next, growth differences (ΔOD_{600}) between phage-treated and untreated controls start to diminish likely due to the emergence of resistant bacteria. Therefore, the **time of bacterial regrowth** (t_{res}) is represented by the first negative valley of ΔOD_{600} followed by a significant increase in ΔOD_{600}. This valley was obtained with the function **find_peaks** (window size: 5, threshold: 0.05) of the R package **ggpmisc** [273,274]. If an increase of ΔOD_{600} >=0.05 was not observed after phage lysis, it was considered that no significant resistance had occurred in the time window analyzed. In that cases, the last time point of the time course was assigned as $t_{res.}$

Finally, the **percentage of phage inhibition (pIAUC)** was determined as follows: $pIAUC=1-(AUC_{infected}/AUC_{non\text{-}infected})$[269], being AUC the area under the curve. AUC values were obtained with the auc() function of the **pROC R** package which uses the trapezoid rule [275].

4. Results

4.1 Kpn-Phage, a collection to study phage-bacteria interactions

Our first goal was to create a representative collection of bacteria and phages to thoughtfully study the determinants of phage host-range in *K. pneumoniae*. Genomic surveillance sequences represent a great opportunity to select strains and conduct hypothesis-driven studies. Here, we have taken advantage of an epidemiological surveillance program for *K. pneumoniae* in the Comunitat Valenciana (Spain) to study phage host range. We tried to maximize the capsular diversity to shed light on the specific interactions between phages and capsules. In parallel, new *Klebsiella* phages isolated from the environment were sequenced and characterized.

4.1.1 *K. pneumoniae* collection

At the time of the analysis (October, 2019), 1,144 clinical *K. pneumoniae* isolates had been sequenced in our laboratory as part of the Surveillance program of Antimicrobial Resistance (NLSAR) [208] and the Infect-ERA consortium [209](**Table S6**). All included samples were from hospitals of the Comunidad Valenciana region, Spain. Sequences were assembled *de novo*, and the number of contigs and N50 values were obtained as metrics of assembly quality [276]. We also determined the capsular locus type (CLT), O-antigen locus type (OLT), and sequence type (ST) of each strain (**§3.3.1**). Excluding sequences with 'None' or 'Low' confidence scores, we observed 73 distinct CLTs for the remaining 1,030 sequences, with 53 CLTs having more than one bacterial strain. Of those 53 CLTs, 55% had more than one OLT, up to 4 (median: 2). The majority of CLTs (79%) were also represented by more than one ST, with a maximum of 10 (median: 2)(**Table S6**). Based on this, we requested 201 strains to the hospitals that maximized both the number of CLTs and the diversity within each capsular type. We prioritized those strains with better assembly quality, that is, with a lower number of contigs and higher N50 values.

Of the requested strains, 29 (14%) were missing or were unable to grow after several attempts. For the remaining 172 strains, the *wzi* sequence (a conserved but hypervariable gene of the capsule operon) was obtained by Sanger sequencing. To confirm that the strain collected corresponded to the strain originally sequenced by the NLSAR, the Sanger *wzi* sequences were compared to the *wzi* of previous Illumina assemblies. We confirmed the identity of 80% of the strains, retaining 138 hosts for the Kpn-Phage collection (**Table S2**). The resulting bacteria included 59 different CLTs, 33 of them with 2 to 7 distinct strains. The collection also included 14 OLTs and 78 STs (**Figure 4.1** and **Table S2**). In the maximum-likelihood phylogeny of the core genome of *K. pneumoniae* strains, 5 isolates showed very long branches (CU371, KP42, X22, KP205, and HGUC_552). These strains obscured the evolutionary relationships among the rest of the samples, which formed a well-supported phylogeny (**Figure 4.1**). Specifically, strains CU371 and KP42 were identified as *K. quasipneumoniae* and strain X22 as *K. variicola*. Kleborate classified strains KP205 and HGUC_552 as *K. pneumoniae*, but with 'Very Low' confidence (**§3.3.1**). Therefore, these 5 strains do not belong to *K. pneumoniae sensu stricto*, but were kept in the collection as outgroups (**Table S2**).

ST11 was the most represented in the collection (n=12), followed by ST15 (n=9) and ST437 (n=7). These STs fall into clonal groups (CGs) 258 and CG14/15, associated with nosocomial spread of multidrug-resistant *K. pneumoniae* [277]. On average, the isolates encoded 6 resistance genes (**Figure 4.1** and **Table S2**). Concretely, 59 strains had a resistance score of 0 (no extended-spectrum beta-lactamases or carbapenemase gene), 60 strains had a resistance score of 1 (presence of an extended-spectrum beta-lactamase) and 19 strains had a resistance score of 2 (presence of a carbapenemase)(**§3.3.1**). There were no strains resistant to both colistin and carbapenems (resistance score of 3)(**Table S2**). This could help with the manageability of this collection in laboratory settings, together with the fact that 132 out of the 138 strains had virulence scores below 2 (no virulence factor or only yersiniabactin).

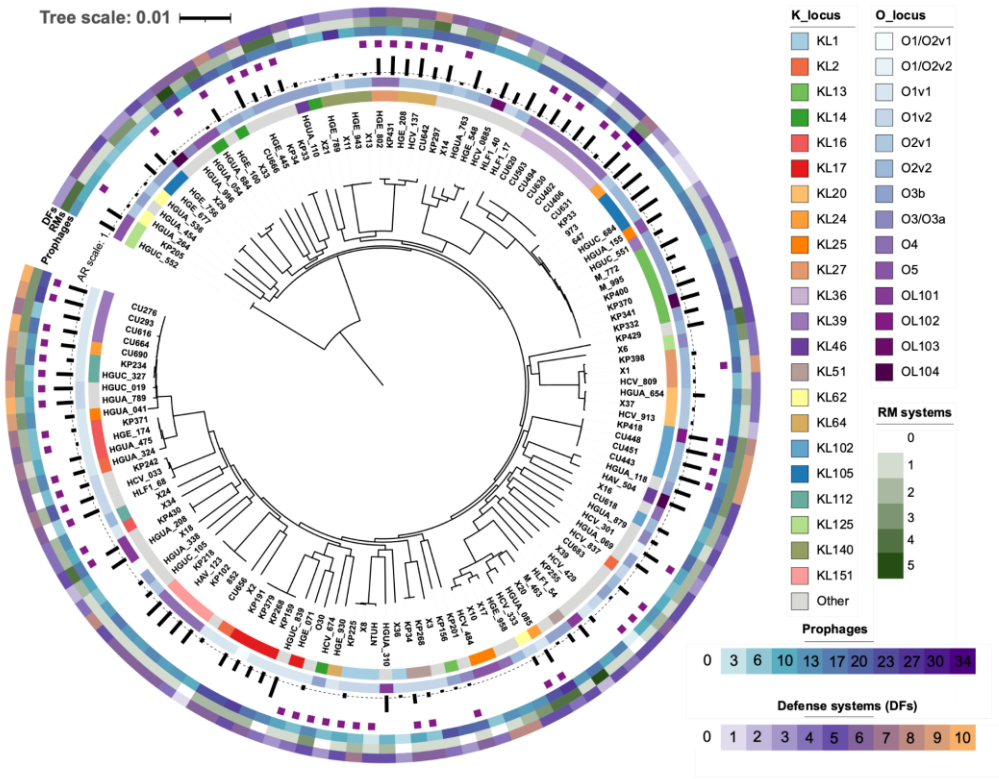

Figure 4.1. Core genome maximum-likelihood phylogeny of *K. pneumoniae* strains included in the collection (§3.3.2). The innermost ring shows the CLT(K-antigen prediction, K_locus) followed by the OLT (O-antigen prediction, O_locus). CLTs with fewer than three strains are shown as 'Other'. Black bars indicate the number of resistance genes to different antibiotic classes. Purple squares indicate the presence of CRISPR-Cas loci. The number of prophages, restriction-modification systems (RMs), and other defense systems (DFs) are also shown for each strain. Strains that did not belong to *K. pneumoniae sensu stricto* (n=5) were omitted from this view as the resolution of the tree was lost. Most nodes (80%) were well-supported (bootstrap values ≥ 95%). Tree scale: nucleotide substitutions per site.

Additionally, bacteria differed in their putative phage defense mechanisms (**Figure 4.1** and **Table S2**). Specifically, 57 strains (41%) encoded at least one CRISPR-Cas system. This prevalence was higher than that reported in previous studies of clinical *K. pneumoniae* isolates, which found 21% and 31% CRISPR-Cas-positive strains [278,279]. As for prophages, their number ranged from 2 to 34 (median: 14). Although these values are at least double than previous estimates [173], this can be due to the use of PhageBoost, which maximizes the recall rate [280]. All the strains had at least two restriction-modification systems (RMs) with a maximum of 15 per strain (median: 7). Furthermore, we found 42 anti-phage systems from the catalogue of 59 additional systems tested with defense-finder [98]. Each defense system was present in 1 to 128 of the 138 bacterial strains, with a median prevalence of 4.71%. Defense systems present in a single strain included Gao, PARIS, RexAB, and Rst. On the contrary, the most prevalent defense was the AbiEii abortive infection system, followed by Retron, CBASS, and Gabija systems, all of which were found in more than 25% of the strains (**Table S2**). This anti-phage repertoire agrees with the results of the complete *Klebsiella* RefSeq database [98].

Finally, we determined whether the bacteria included in the collection were representative of the overall diversity of *K. pneumoniae*. To do so, we first extracted the representative sequence of each CG of *K. pneumoniae* [198] and compared them to the genomes of our strain collection (**§3.3.2**). We observed that the selected bacteria were scattered throughout the whole phylogeny of *K. pneumoniae* CGs, therefore being representative of the species diversity (**Figure 4.2**).

In summary, the bacterial strains selected for the Kpn-Phage collection were variable and representative of the species in terms of their capsule composition, lineages, antibiotic resistance, and phage defense mechanisms.

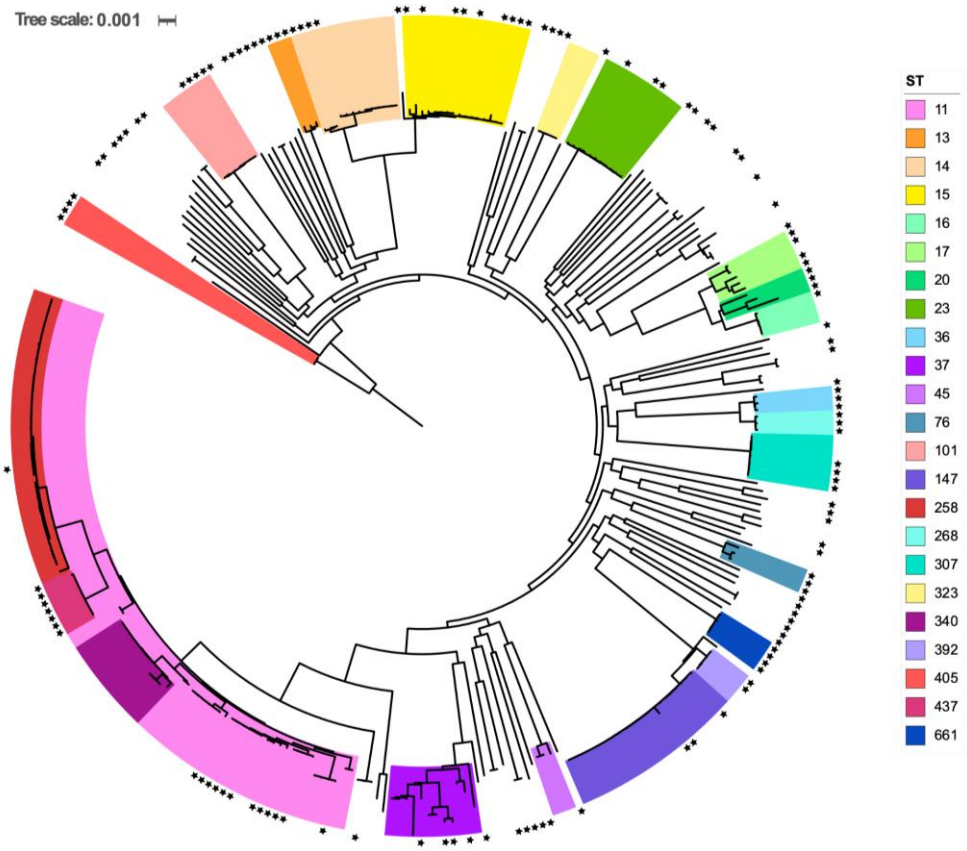

Figure 4.2. Core genome maximum-likelihood phylogeny of *K. pneumoniae* strains included in the collection and representative sequences of CGs (§3.3.2). Sequences of this work are marked with asterisks. Colors represent major sequence types (STs). Strains that do not belong to *K. pneumoniae sensu stricto* (n=5) were omitted from this view. Tree scale: nucleotide substitutions per site.

4.1.2 Phage sequencing

For phages, 71 isolates provided by the Environmental and Biomedical Virology Lab (I2SysBio, UV-CSIC) were purified, sequenced and characterized. These were isolated from water and soil samples near a sewage water plant in Valencia (Spain). For this, 36 clinical *K. pneumoniae* strains available at that time, hereinafter referred to as isolation hosts, were used. Isolation hosts included 19 bacterial strains from the aforementioned Kpn collection, but also bacteria whose genomic sequence was not available. For the latter, their capsular type was inferred by the *wzi*-typing method (§3.2.3). In total, 26 capsular types were used for phage isolation (Table S2).

Prior to sequencing, phages were amplified in their corresponding host and purified. We chose ultrafiltration as the preferred method for purification because it produced purified phage with comparable titers and less host DNA than PEG (results not shown). Phage titers ranged from 2.50E+08 PFU/mL to 8.00E+11 PFU/mL (median: 2.50E+10) and resulting DNA ranged from 0.19 ng/µL to 25 ng/µL (median: 1.26 ng/µL). We found a weak correlation between DNA yield and input phage titer (Pearson's r=0.17, p=0.08), suggesting that other phage characteristics affected the extraction process.

Sequencing was successful for all except one phage, for which only 373 reads were obtained. This was not due to the use of a low viral titer as input (1.25E+10 PFU/mL) or failed DNA extraction (0.767 ng/µL). For the remaining 70 viruses, the sequencing run yielded between 36,819 and 520,066 paired-end reads (median: 167,318). Bacterial reads were less than 10% for 65 out of the 70 phages, confirming successful phage purification and host DNA removal (**Table S3**). Sequence depth ranged from 317X to 2,791X (median: 1,443X) which allowed obtaining a single contig for all viruses after using different *de novo* assembly approaches (**§3.5.1**). All but nine phage genomes could be circularized, indicating complete phage genomes (**Table S3**).

4.1.3. Clustering of phages

To identify duplicate viruses, we first compared all the genomes with fastANI, and clusters of highly similar phages were observed (**Figure 4.3**). A threshold of 99% in ANI and coverage was set to consider phages as duplicates and one representative of each cluster was retained for further analysis. This resulted in 46 phage strains (**Figure 4.4**). Of the 17 phages with biological duplicates, most were isolated with the same bacteria/capsular type (13 and 15, respectively) or from the same environmental sample (12). However, highly divergent phages were also isolated from the same sample location and bacterial host (**Figure 4.4**).

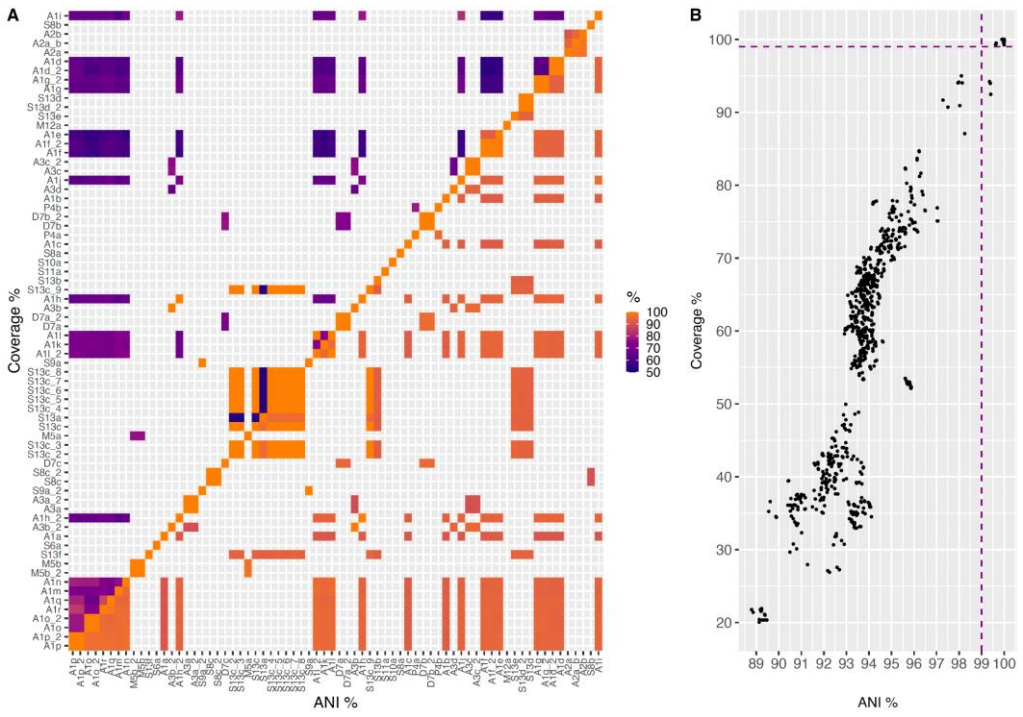

Figure 4.3. Identity between phage genomes and selection of representative phage strains (§3.5.2). A. Pairwise comparison between sequenced phages with fastANI (fragment length=50 nt). The upper triangle shows coverage values (%) and the lower triangle average nucleotide identity (ANI) values (%). **B.** Relationship between ANI and coverage values. The purple dashed line marks the threshold used to determine independently isolated plaques of the same phage strain.

After genome deduplication, phage strains were classified into different groups and species using intergenomic similarity (IGS) thresholds of 45% and 95%, respectively. We applied a 45% threshold instead of the current 70% to defining genus as the latter was published after the analyses were performed [21]. Furthermore, this would lead to a high number of discrete groups with unclassified viruses. To avoid confusion, we will refer to them as similarity groups instead of genera. We obtained 13 similarity groups (> 45% IGS), with groups 6 and 9-12 composed of singleton phages. Phages were named with a first letter, indicative of the phage family (§4.1.5), a number indicative of the phage group, and a letter indicative of each species. In terms of species, only phages A2a and A2a_b were highly related (> 95% IGS), resulting in 45 different species out of the 46 phage strains (Figure 4.4).

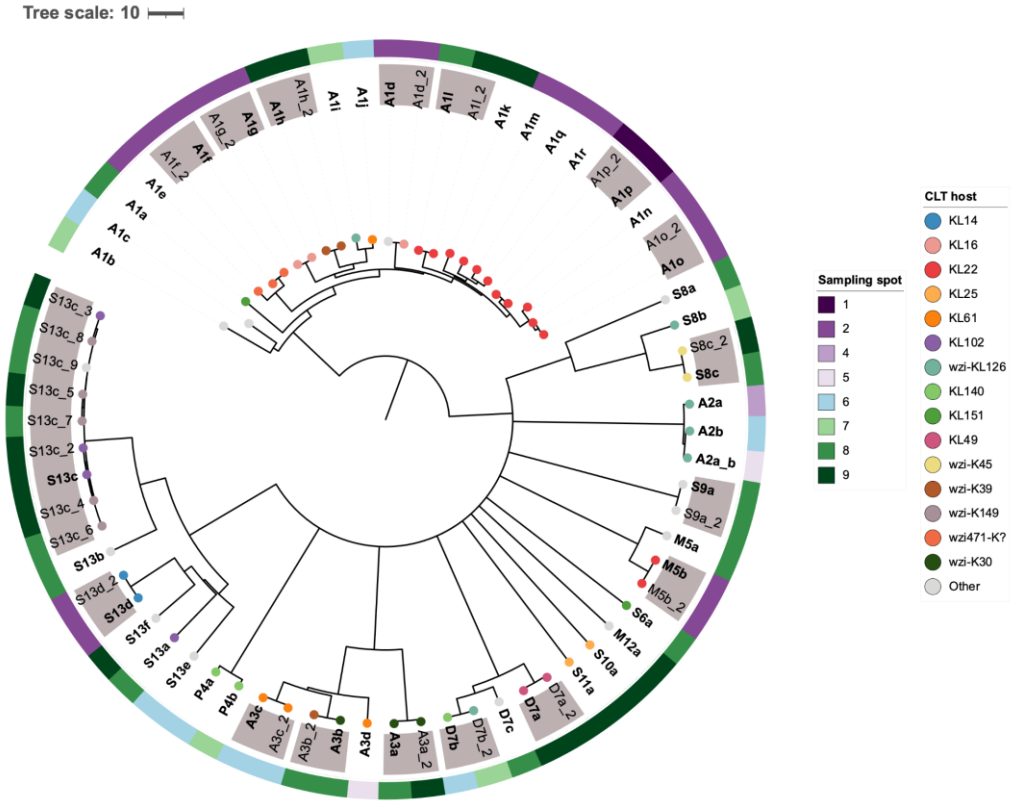

Figure 4.4. Dendrogram of the 70 phage isolates (§3.5.2). Neighbor-joining tree obtained from ANI distances for the sequenced phage isolates. Phages that were considered duplicates are marked with gray boxes, and the selected representative of each phage strain is indicated in bold. This resulted in 46 distinct phages that were selected for further study. A2a and A2a_b represented the same viral species. The capsular type of the *Klebsiella* host used for phage isolation is indicated with colored circles at the tips of the dendrogram. Capsular types with one occurrence are shown as 'Other'. ?: indicates *wzi* alleles whose correspondence with a CLT is unknown. The different sampling spots from which phages were isolated are indicated with colors outside the dendrogram (§3.1.2). Tree scale denotes ANI distance (100-ANI).

4.1.4 General characteristics of phages

Phage plaque morphologies varied among the 46 phage representatives in size, opacity, and presence of haloes suggestive of phage depolymerase activity (Figure 4.5 and Table S7). Overall, phages exhibited clear plaques except D7a,

which showed plaques with turbid morphology, indicative of putative temperate behavior. Plaque size at 24 hours-post-infection ranged from 0.18 mm to 5.47 mm (median: 0.81). Phage plaques were surrounded by diffusion haloes in 35 phages. Depolymerization activity inferred by halo size at 24 h ranged from 0.23 mm to 8.75 mm (median: 1.03). Phages with larger plaque and halo sizes belonged to groups 1 and 3. In contrast, phages with smaller plaques and halo sizes were classified into groups 11 and 13. Representative TEM images of the 13 phage groups were obtained, which allowed classification of phages into different morphotypes (myovirus, siphovirus, and podovirus) (Figure 4.6).

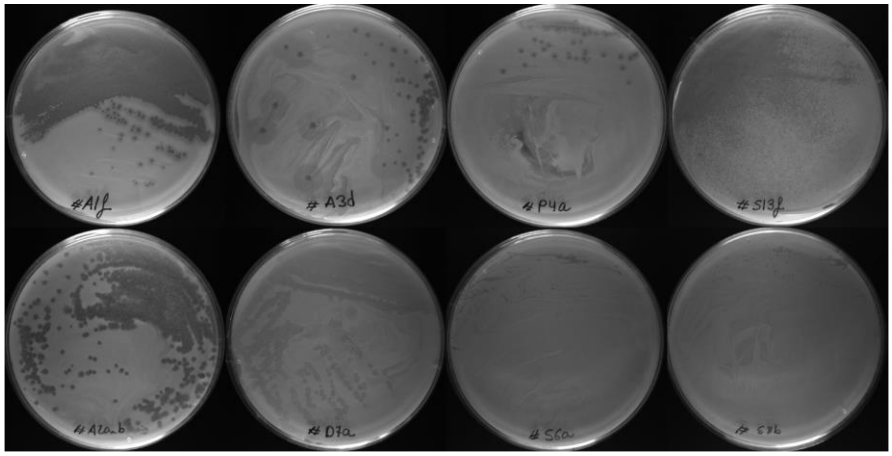

Figure 4.5 Plaque morphology for a subset of phages (§3.4.4). In the top, phages with plaques surrounded by haloes are shown. In the bottom, phages with plaques without haloes are shown.

In terms of genomic characteristics, the collection included diverse dsDNA-tailed viruses, with sizes ranging from 39,664 bp to 144,995 bp (median: 45,022 bp)(Figure 4.7). Genomic GC content ranged from 44% to 57%, with a median of 48%. Interestingly, 9 phages coded for tRNAs, with phages from group 5, M5a and M5b, having 22 and 23 tRNAs, respectively (Table S7). The fraction of proteins with unassigned function (i.e., hypothetical proteins) ranged from 0.13 in phage A1e to 0.55 in phage S13f (mean: 0.34; median: 0.33) with the custom annotation pipeline (§3.5.3). When the PHROG approach was used, the number of

hypothetical proteins was considerably higher (mean: 0.53, median: 0.52). To avoid propagation of errors in the databases, we kept the PHROG annotation, which is now the gold standard [250](**Figure 4.7**). Still, results from the custom pipeline pointed out that the PHROG approach might be too conservative and that further work might decrease the number of unannotated phage proteins. Interestingly, some phages presented immunity-related proteins, such as anti-restriction (n=3) or antitoxin proteins (n=1) (**Table S7**).

Regarding phage lifestyle, all phages except those in groups 5 and 9 were predicted to be lytic based on the absence of integrases. However, some phages might present a facultative lytic lifestyle given that they presented *parAB* genes (n=5), involved in plasmid-like replication, or transcriptional repressors (n=13), involved in the control of the lytic-lysogenic cycle. Interestingly, the results of the *in silico* prediction of phage lifestyle with 3 different programs, fully agreed on only 21 phages (46%) (**Table S7**). A consensus lytic lifestyle (>=2 methods) could be assigned to 36 of the viruses. Of note, BACHPLIP assigned a lytic lifestyle to all phages with confidence intervals of 77.5% to 100% (median: 96.59). For those in which a consensus temperate lifestyle was obtained, BACHPLIP confidence scores were high but considerably lower (median: 92.52) than for phages with a consensus lytic lifestyle (median: 97.03). PhageAI assigned a temperate lifestyle for 13 phages, with confidence intervals from 60.4 to 99.97 (median: 98.78). Finally, PHACTS was the program that classified more phages as temperate (n=23). However, 39 predictions from this program had an average probability score less than two standard deviations away from the score of the alternative lifestyle, being unreliable [254] (**Table S7**). Phages had no genes related to antimicrobial resistance or virulence. The only exception was phage M5b which carries a gene with similarity to the lambda *bor* gene, related to bacterial resistance to serum complement killing [281].

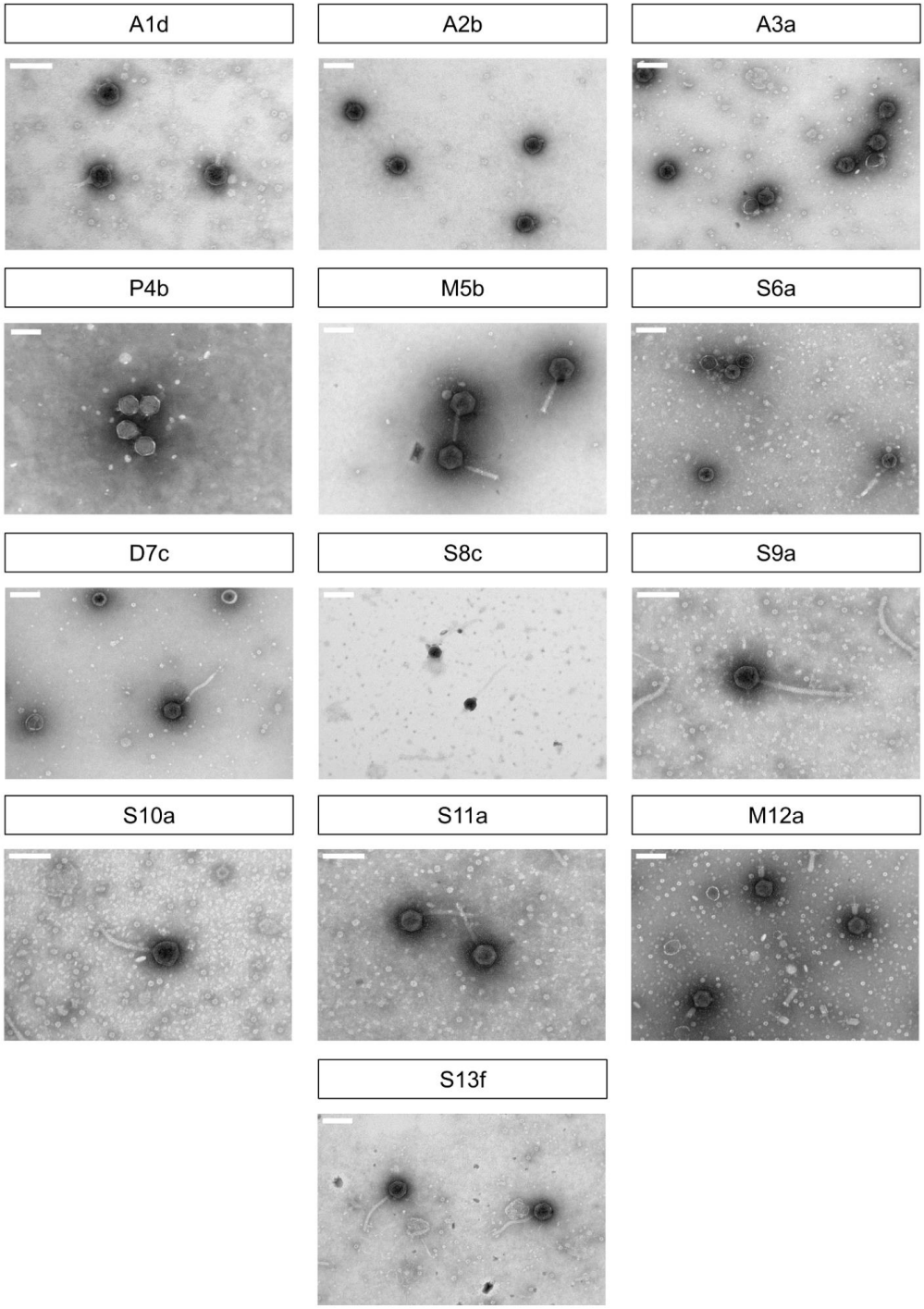

Figure 4.6. TEM of representative phages from each group (§3.4.4). Scale bar for TEM: 100 nm.

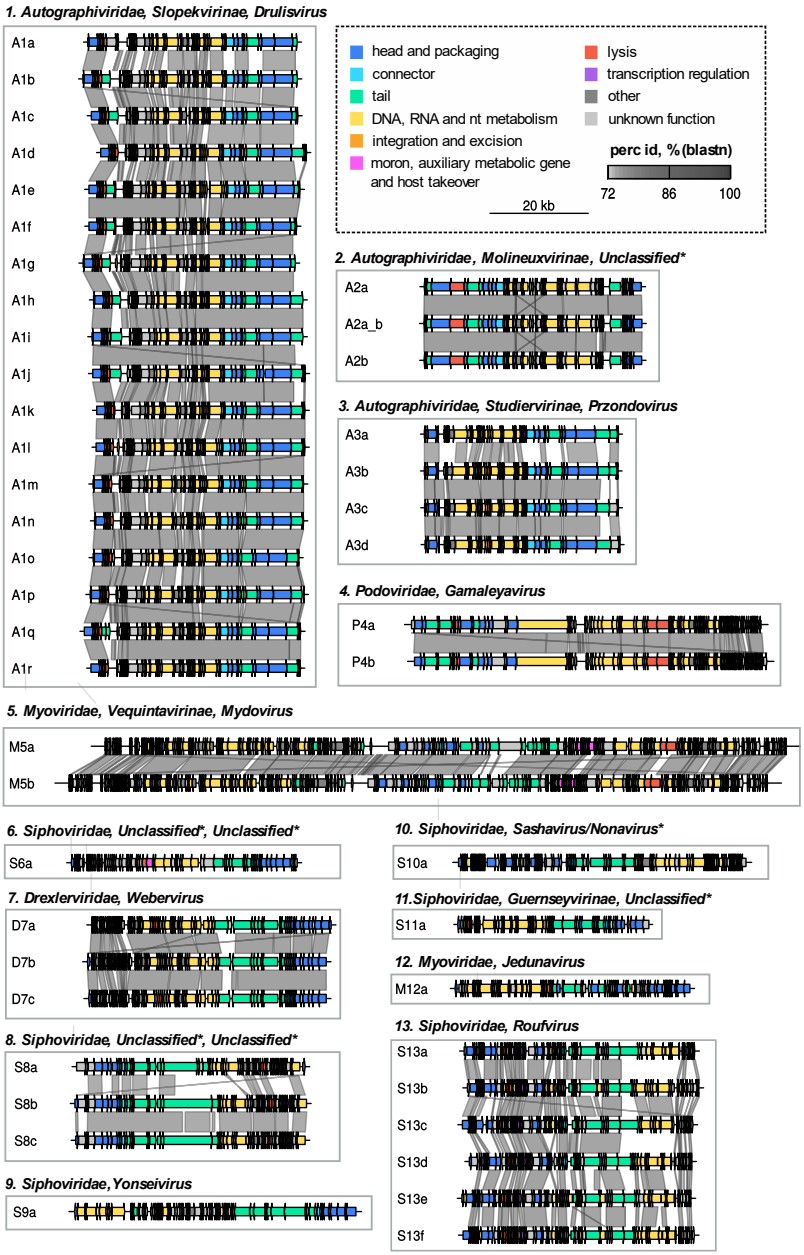

Figure 4.7. Genomic comparisons and functional annotation of *Klebsiella* phages of the Kpn-Phage collection (§3.5.2-3). The 13 phage similarity groups are shown as well as their closest viral family, subfamily (if available), and genus after comparison with database sequences. A: *Autographiviridae*; D: *Drexlerviridae*, M: *Myoviridae*; *P: Podoviridae*; S: *Siphoviridae*. Phage genomes were arbitrarily permuted at the terminase small/large subunit for easier visualization of synteny after blastn comparison. Arrows represent ORFs and are colored based on functional categories of PHROGS. Phage groups with asterisks denote incomplete taxonomic classification.

4.1.5 Phage classification

To classify the new phages, we compared them with phages from the databases with available taxonomic classification. Comparison with phage sequences from GenBank (February 2021, n=14,037), resulted in 478 genomes showing at least 30% coverage and 70% nucleotide identity with at least one of the 46 phage strains. Of the 478 genomes, 38.7% had been isolated in *Klebsiella* sp., followed by phages isolated in *Escherichia* sp. (12.8%) and *Salmonella* sp. (11.9%). Next, pairwise IGS comparisons between all genomes (478+46) were used to define similarity clusters, retaining 204 phage genomes that were highly related to our phages (IGS > 45%). Of these, 89.7% were *Klebsiella* phages. Only one phage (S8b) showed > 95% IGS with previously described sequences (*Klebsiella* phage vB_Kp3), suggesting that the rest represent new species (**Table S7**). Also, group 2 was the only monophyletic group when compared to all phage NCBI genomes with ViPTreeGen (data not shown). Therefore, despite phages being isolated from a nearby geographic location, diversity was seen within groups except for A2 phages (**Figure 4.7** and **4.8**). Phages in groups 2, 6, and 8–11 were related to less than 10 database phage sequences, showing that these groups were underrepresented. On the other hand, phages from groups 1, 3, and 7 clustered with many phage genomes available (n > 50) (**Figure 4.8**).

Based on pairwise IGS values, we found that new phages were distributed in five families of the order Caudovirales: *Autographiviridae* (n=25), *Drexlerviridae* (n=3), *Myoviridae* (n=3), *Siphoviridae* (n=13), and *Podoviridae* (n=2)(**Figure 4.7**). This classification agreed with the phage morphotypes observed by TEM (**Figure 4.6**). In order to rule out the possibility that new related phages had been deposited and classified since the initial analysis, phages that were not classified at the subfamily and/or genus levels were again confronted against the NCBI (Accessed: 01/03/22). After this, phages from groups 1, 3, 4, 5, 7, 9, 12, and 13 were classified within *Drulisvirus*, *Przondovirus*, *Gamaleyavirus*, *Mydovirus*, *Webervirus*, *Yonseivirus*, *Jedunavirus*, and *Roufvirus*, respectively. The remaining groups could not be assigned at the genus level.

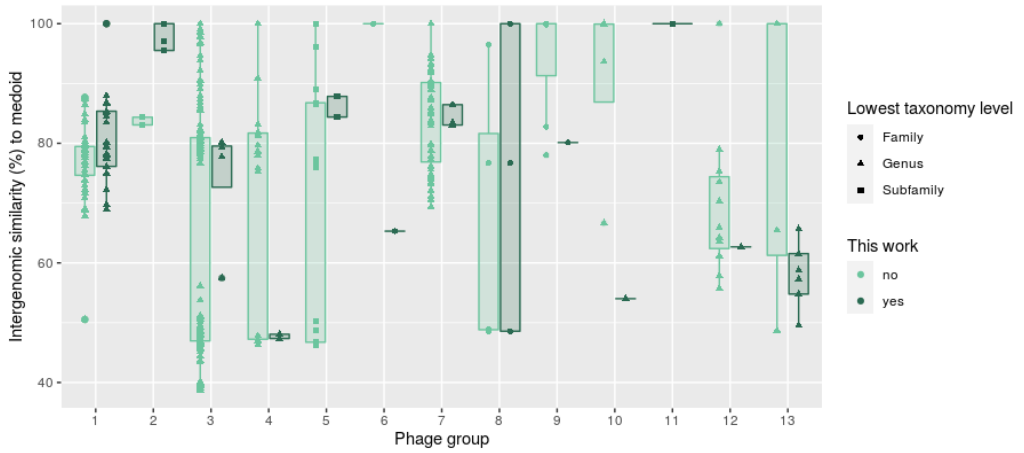

Figure 4.8. Diversity within phage groups including GenBank genomes (§3.5.2). Groups were determined based on clusters of ≥ 45% intergenomic similarity values (IGS) with a medoid sequence. Dots represent pairwise IGS values to the medoid genome of each group. The shape of points indicates the lowest taxonomy available for each phage group based on ICTV ranks (February, 2021). Classification of phages of groups 5 and 9 was unavailable at the time of this analysis. Recently, they were classified as *Mydovirus* and *Yonseivirus*, respectively.

To decipher the relationships of unclassified phages at lower taxonomic ranks, we further analyzed phages in groups 2, 6, 8, 10, and 11 (Figure 4.9). Specifically, we found that A2 phages could represent a new genus, which includes *Proteus* phage PmP19 (MT680615) and *Klebsiella* phage vB_KpP_FBKp16 (MW394389), as their IGS values are > 70 % (Figure 4.9.A). Phage S6a is an orphan phage related at the subfamily level with *Salmonella* phage PMBT28 (MG641885) (IGS = 65.3 %) (Figure 4.9.B). S8 phages belong to the same subfamily (IGS > 45 %) but to different genera: phage S8a belongs to the same genus as *Enterobacter* phages ATCEA85 (MN656993) and ATCEA23 (MW419910), while phages S8b and S8c are grouped with *Klebsiella* phages vB_Kp3 (same species as phage S8b) and vB_KleS-HSE3 (MT075871) (Figure 4.9.C). The only representative of group 10, phage S10a, is related at the subfamily level to both *Sashavirus* and *Nonanavirus* phages (IGS > 45 %), but represents a new genus (Figure 4.9.D). Finally, the phage in group 11, S11a, is an orphan phage within the subfamily *Guernseyvirinae*, representing a new genus (Figure 4.9.E). Therefore, from the analyses of the 5 unclassified groups, we suggest that they represent 3 new subfamilies and 5 new genera.

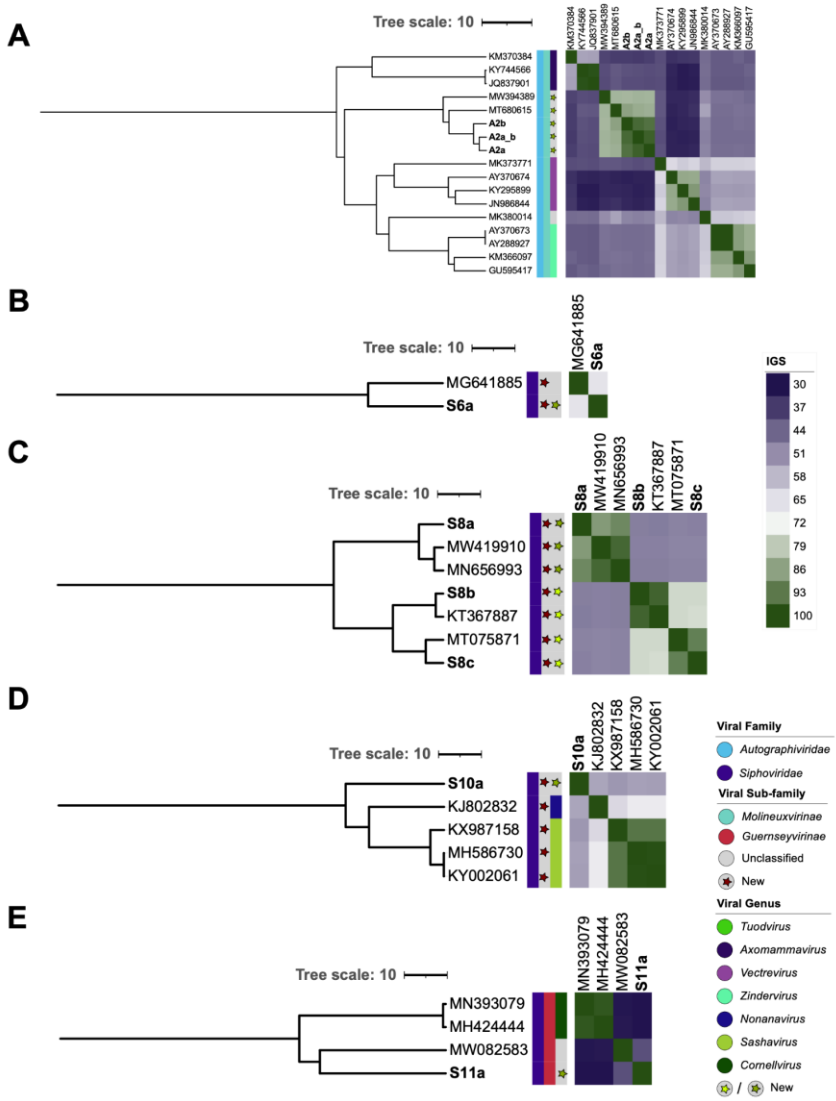

Figure 4.9. Dendrograms of unclassified viruses based on IGS values (§3.3.2). Phages that were not classified at the subfamily and/or genus levels by similarity (IGS<45%) with any phage from the database were confronted against the NCBI (Accessed: 01/03/22) to exclude that new related phages had been deposited and classified since the initial analysis. Colored strips denote family, subfamily, and viral genus. New subfamily and genera suggestions are marked with a star and are based on updated ICTV thresholds [21]. Phages from this work are in bold. **A.** Phages related to group 2. **B.** Phages related to S6a phage. **C**. Phages related to group 8. **D.** Phages related to S10a phage. **E.** Phages related to group 11. The tree scale denotes IGS distance (100-IGS).

4.1.6 Contextualization of phages with previously reported *Klebsiella* phages

In the previous sections, we characterized the diversity of the phage collection. However, until this point, we did not know whether this is representative of the overall known diversity of *Klebsiella* phages. To address this, we obtained all *Klebsiella* phage genomes available (February 2021, n=318) and calculated genome-wide proteomic similarities. Phages of the collection were spread among the main *Klebsiella* phages families, with only minor families, *Ackermanviridae (n=21)* and *Demerecviridae (n=9)*, not being represented. Phages from common subfamilies *Tevenviridae* (n=37) and *Peduovirinae* (n=17), were not recovered. Apart from these exceptions, the known diversity of *Klebsiella* viruses was covered in this collection as Kpn-phages were scattered throughout the whole proteome tree of these viruses (**Figure 4.10**).

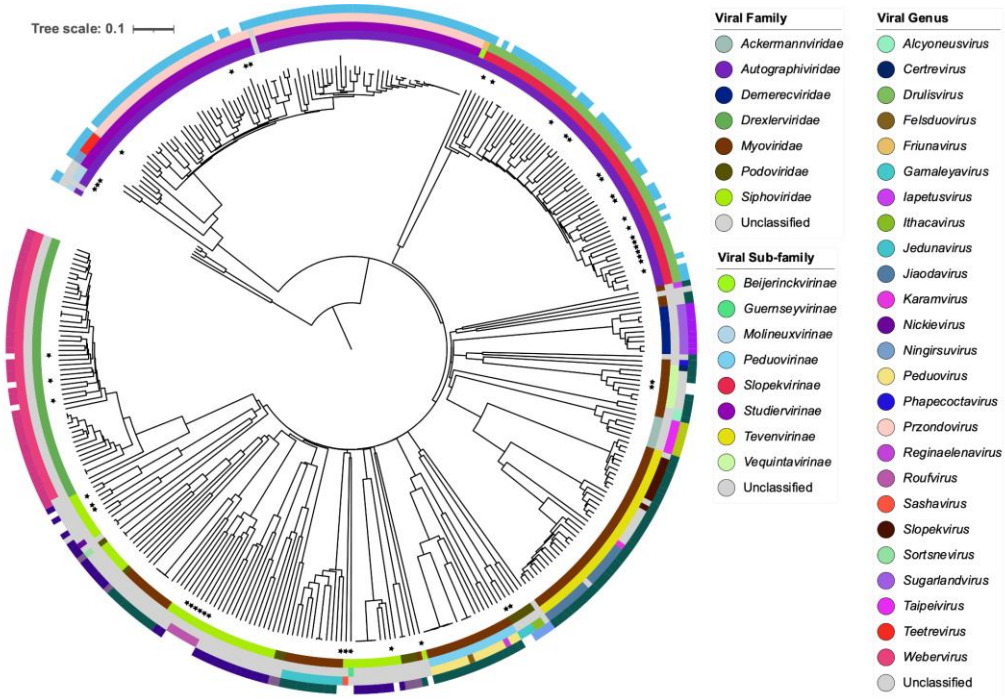

Figure 4.10. Diversity of phages included in the Kpn-Phage collection compared to previously described *Klebsiella* viruses (§3.3.2). Proteomic tree of all *Klebsiella* phages (February 2021), including the 46 newly sequenced phages (marked with a star) after computing genome-wide sequence similarities with VipTreeGen. The tree scale shows global genomic distances. Clades of phages were colored based on the ICTV taxonomy.

NOTE: *A few days before the submission of this work, an official proposal to the ICTV was submitted led by Dr. Andrew Kropinski. In this proposal, the members of the S13 similarity group were proposed as a new subfamily, Fartonvirinae. This name was based on the tail morphology of these siphoviruses which reminds of a traditional Valencian pastry known as "fartón". Besides, each member of this group was proposed as a new genus. Concretely, phage S13a as a member of Hortavirus, phage S13b as a member of Lapuntavirus, phage S13c as a member of Levantevirus, phage S13d as a member of Alboraiavirus, phage S13e as a member of Acequiavirus and phage S13f as a member of Turiavirus. Each name refers to Valencian places, related to the isolation of phages. Additionally, phage M12a was suggested as a new genus named Mascletvirus. This name was given based on a type of firecracker used in Valencia and called "masclet", whose shape somewhat resembles the virus. Phages of group P4 have been reclassified into Kaypoctavirus with a new Klebsiella phage KP8 (https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?mode=Tree&id=2842786&lvl=3&p=7&lin=f&keep=1&srchmode=1&unlock). They are related to Gamaleyavirus at the subfamily level (Enquatrovirinae). Phages S6a, S8a-c and S10a remain unclassified at the subfamily level. Phage S11a remains unclassified at the genus level. Updated and proposed taxonomy to the ICTV was added to* ***Table S7*** *(Accessed: 17/02/23).*

4.2 Genetic determinants of host tropism in *Klebsiella* phages

In the previous section, we found that the Kpn-Phage collection was diverse and representative of the known diversity of *K. pneumoniae* and its phages. In this part, we aimed at quantifying the determinants of host-range in this species, with special emphasis on the role of capsule diversity and phage depolymerases. As several strains with the same capsular type were included, this allowed us to determine strain-specific differences. We performed all-vs-all pairwise comparisons by spot-tests and confirmed these results by bactericidal, progeny, depolymerization, and adsorption assays. This information was combined with the genomic analyses of phages and bacteria to decipher the genetic basis of phage tropism.

4.2.1 Host tropism of isolated *Klebsiella* phages

We first tested the ability of phages to produce a clearing on the bacterial overlays of the 138 putative hosts. We performed triplicate spot tests for all the 71 x 138 = 9,798 phage-bacteria pairs, with 6,348 being unique after deduplication of phage strains (§4.1.3). Additionally, we evaluated 1,242 phage-bacteria interactions with two phage concentrations (9 phages, against all the 138 strains, Table S7). Also, 136 phage-bacteria pairs of interest were re-tested by spotting 8 serial dilutions of phages in duplicate. We scored an interaction as positive if at least two thirds of the replicates showed a phage effect (single plaques, turbid or clear spot) in the bacterial lawn. To simplify, we will only refer to the results obtained with the 46 phage representatives (Figure 4.4) hereinafter.

The concentrations of the 46 phages ranged from 10^4 PFU/spot to 10^7 PFU/spot. However, 80% of the phages were tested with concentrations of 10^5 PFU/spot (n=20) or 10^6 PFU/spot (n=17), that is, only 1 log of difference (Table S7). Spot tests showed 98.9% reproducibility, producing the same outcome (clearing *vs.* no clearing) in the three replicates, with S8 phages showing the highest variability (93%). Additionally, we only observed 4 cases (0.003%) of discrepancies for phages tested with two concentrations.

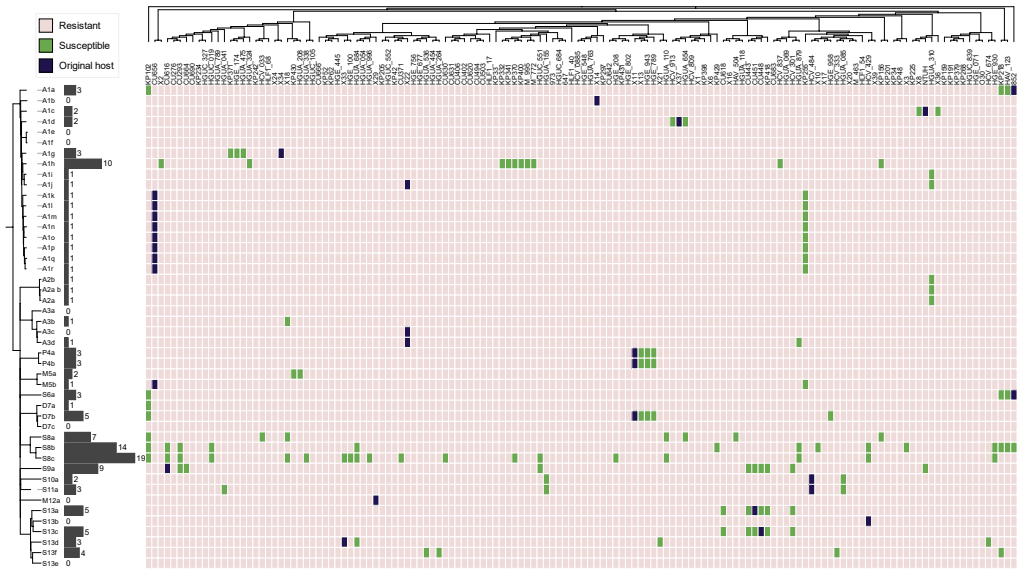

Figure 4.11. Kpn-phage infection matrix. Each row represents a phage representative strain (n=46) and each column a bacterial strain (n=138). For phages, the dendrogram constructed from IGS values is shown, whereas the maximum-likelihood core phylogeny is shown for bacteria. A given bacterial strain was considered to be susceptible to a given phage (green boxes) if at least two-thirds of replicates of the spot assay were positive. Original host-phage pairs in which each phage was primarily isolated are indicated in purple. The black histogram and associated numbers indicate the host breadth of each phage in the 138 total bacterial strains analyzed excluding isolation hosts.

Overall, we found that from the 6,394 unique interactions, only 153 yielded reproducible spots, which were reduced to 124 after removing phage-bacteria combinations used for phage isolation to avoid sampling bias. This suggests that, on average, a *Klebsiella* phage can infect less than 2% of bacteria apart from its isolation host (Figure 4.11 and Table S8). Host breadth ranged from 0 to 19 of the 138 strains (excluding the isolation strain), with two distinct patterns: most phages (42/46) infected one or a few strains from a few CLTs (median: 1, mean: 1.79; 1-3 CLTs), but the four phages in groups 8 and 9 (S8/S9 phages) exhibited a much broader host range (median: 11.5, mean: 12.25; 4-16 CLTs, Figure 4.11 and Table S8). Phages A1b, A1e, A3c, and D7c were unable to spot in any host of the collection, apart from their isolation strain. Conversely, there were 44 *K. pneumoniae* strains of 21 different CLTs that were not infected by any phage.

4.2.2 Quantifying the importance of bacterial traits shaping the spotting matrix

To investigate the determinants of host tropism, we first focused on the 42 non-S8/S9 phages. We assessed which host features predicted spotting by phages (true positive rate, TPR). For each trait (CLT, OLT, ST and protein receptors), we sampled random reference hosts and the phages' spot outcomes and used this information to predict the ability of phages to spot other bacterial strains sharing the same trait. This revealed that CLT was the host trait that best predicted tropism. Specifically, a phage spotting in a given strain had a 92% probability of spotting also in other strains of the same CLT (TPR = 0.92 ± 0.001 for non-S8/S9 phages; **Figure 4.12.A**). Bacterial ST was another good predictor of host tropism (TPR = 0.60 ± 0.01), albeit it is correlated to CLT in our dataset (Cramér's V ES = 0.9). In contrast, OLT and 45 additional proteins encoding putative phage receptors were very poor predictors of host tropism (TPR = 0.04-0.09). This confirmed that the bacterial capsule is a major determinant of host tropism in *Klebsiella* phages.

Phages infecting encapsulated bacteria can encode depolymerases domains (Dpos) in their receptor-binding proteins (RBPs) that specifically digest capsular polysaccharides to provide access to the cell surface [163]. We found RBPs with putative Dpos in 38 of the 42 non-S8/S9 phage genomes. These were in the same ORFs as tail (n=68) o hypothetical proteins (n=12) and were therefore referred to as tail spike proteins (TSPs) [83]. Most of these TSPs should be functional because plaques were typically surrounded by haloes (**Table S7**), the typical indicator of Dpo activity. Supporting this, quantitation of capsule polysaccharide levels showed a strong reduction (t-test $P < 0.001$) in 14/14 bacterial strains inoculated with halo-producing phages relative to non-inoculated controls (**Figure 4.12.B**). Failure to detect TSPs in phages is common [173,206]. For instance, three of the four non-S8/S9 phages with no detectable TSPs showed haloes and significant activity against the capsule, suggesting depolymerase activity (**Table S9**). However, in 20 phages the number of TSPs encoded was larger than the number of susceptible CLTs (**Table S7**), suggesting inactivity of some TSPs or specificity of action towards CLTs not included in this work.

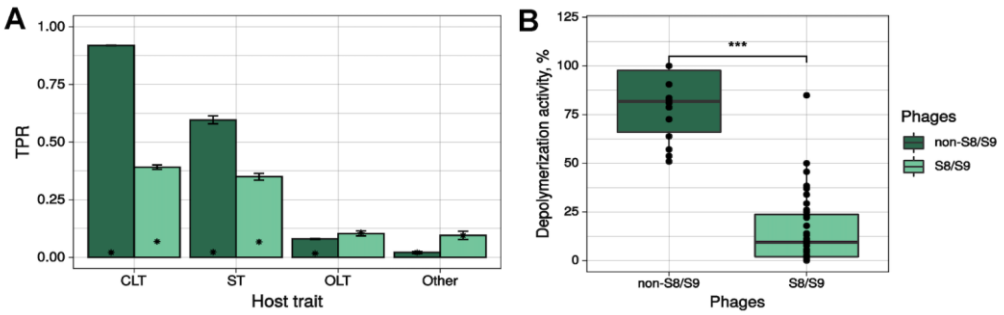

Figure 4.12. **Prediction of phage interactions by bacterial traits. A**. **A**. Sensitivity or True Positive Rate (TPR) of different host traits to predict positive spot tests for non-S8/S9 phages and phages from groups S8/S9. Bars represent the standard error of 3 independent calculations. Asterisks represent the mean TPR values obtained with randomized data (null expectation). CLT: capsular locus type; OLT: O-antigen locus type; ST: sequence type. Other: presence of 45 secondary receptors (pooled). **B.** Depolymerization activity of halo-producing non-S8/S9 phages and non-halo-producing S8/S9 phages relative to non-inoculated controls. Each point represents the mean activity of a given phage-bacterium combination tested. Asterisks indicate statistical significance (t-test: $P < 0.001$).

Inactivity of some TSPs could be the result of the enzymatic domain being truncated or not incorporated into the phage particle. To assess this, we determined the TSPs architecture of each phage following a previous nomenclature [163]:

We found 13 podoviruses with 2 TSPs. Six of them only had a Dpo domain because the first TSP was truncated (Figure 4.13). Additionally, we found 2 Dpos in phages A1l and A1q without the structural domains, suggesting that they could be soluble enzymes (Figure 4.13 and Figure 4.14). Interestingly, some podoviruses encoded adhesins (n=3) and acetyl xylan esterases (n=2), also involved in host recognition [174,282].

All siphoviruses showed one TSP with a Dpo domain (Figure 4.13). As an exception, *Klebsiella* phage D7a presented an esterase domain, which agrees with the lack of plaque haloes (**Table S7**).

Finally, the two myoviruses showed a more complex TSP architecture, in agreement with phages of their genome size (>100 kb) [163]. Even though we detected structural domains typical of TSPs, phage M5a presented 2 Dpos and phage M5b just one (**Table S7**).

In summary, 5 non-S8/S9 phages had no detected Dpos, 28 phages encoded one, 8 phages encoded two domains and the phage A1q encoded three (Figure 3D). This resulted in 27 phages with the expected dosage of Dpos and susceptible CLTs (1:1), 8 phages with less Dpos than susceptible CLTs, and 7 phages with more Dpos than CLTs. Dpos showed the expected secondary structure when modelled with AlphaFold2 (Figure 4.14).

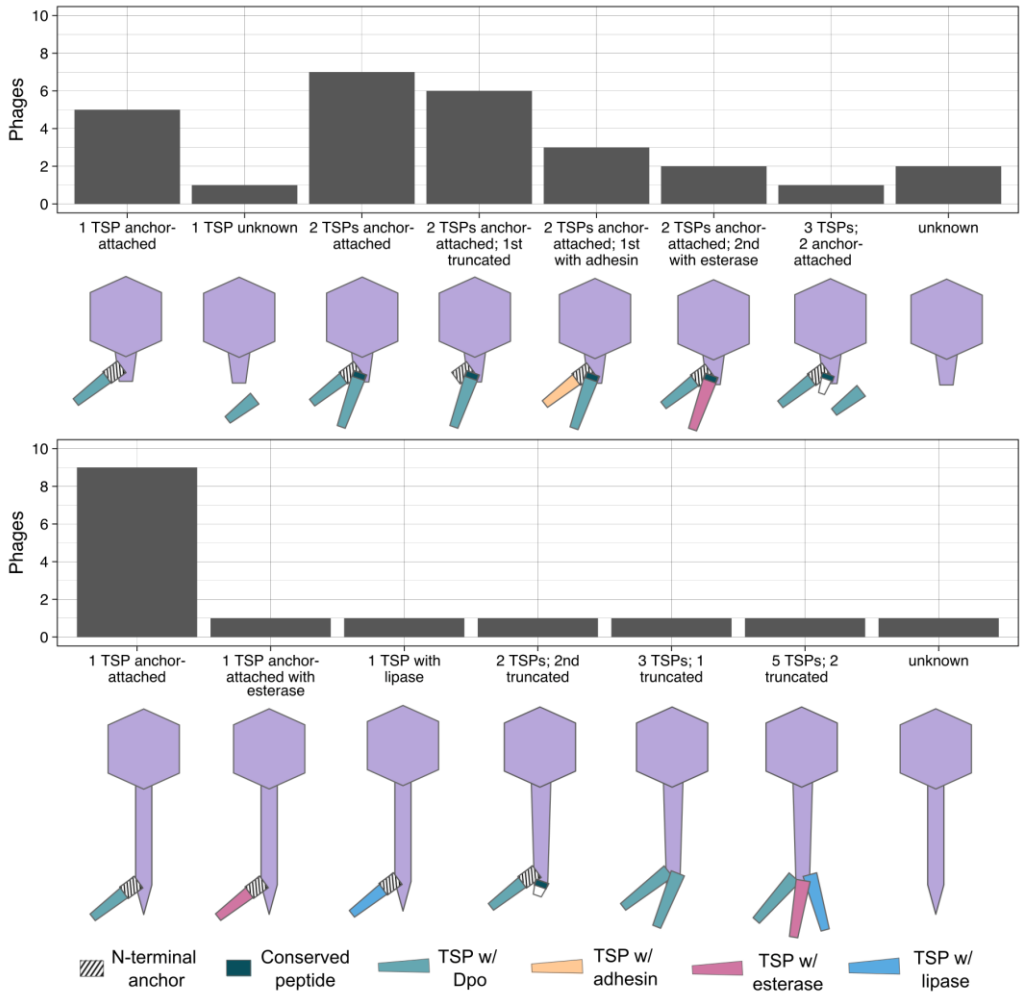

Figure 4.13. Architecture of RBPs of non-S8/S9 phages. Predicted tail spikes (TSPs) with or without depolymerases (Dpos) are shown. The primary TSP is attached via an N-terminal anchor and the secondary TSP via a conserved short peptide. Top: podoviruses. Bottom: Sipho and myoviruses. Only one copy of TSP is shown for simplification.

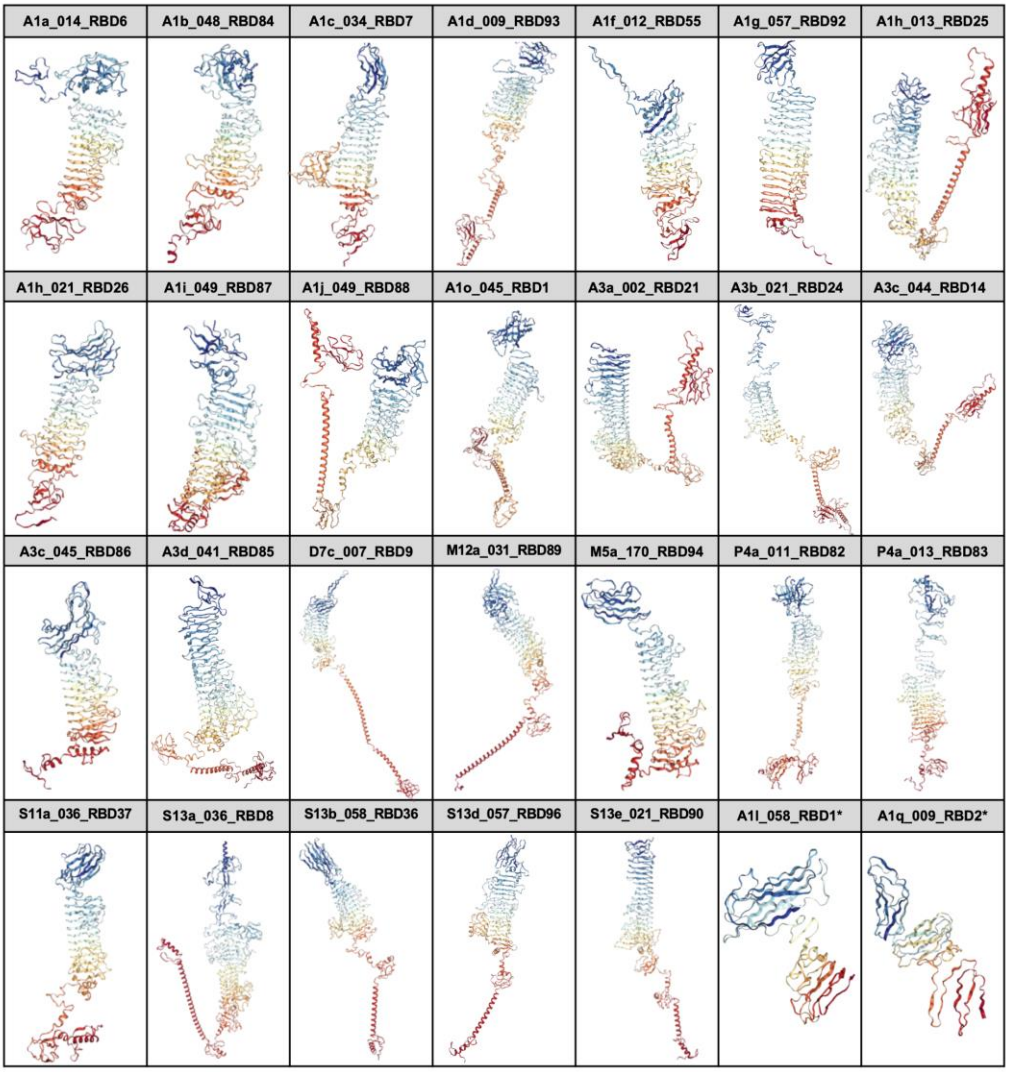

Figure 4.14. Representative tertiary structure of TSPs with predicted Dpos. Each structure is labelled with the phage name, the protein id and the assigned RBD cluster, if applicable. Asterisks denote putative soluble depolymerases.

4.2.3 Sequence-based prediction of phage capsular tropism

The capsular type of bacteria can be obtained from their genome [29] but at present, it is not possible to predict the capsular tropism of phages in a similar way. To address this, we first expanded our dataset by including 64 additional *Klebsiella* phages from the literature with known tropism (**Table S1**). We then focused on the central and C-terminal regions of TSPs with Dpos, referred to as receptor-binding domain (RBD). To avoid the inclusion of truncated TSPs from phages of the literature, we only retained those with at least 500 *aa*. We extracted 46 RBDs from the 42 non-S8/S9 phages and 117 RBDs from additional *Klebsiella* phages. The 163 RBDs grouped into 35 clusters with at least two RBDs sharing > 40% sequence coverage and > 50% amino acid identity (**Table S10**).

RBD cluster thresholds were chosen based on the minimization of the number of clusters and increased sensitivity (results not shown). We found that RBD clusters had a mean TPR of 58.1 ± 4.0% on phage tropism (**Figure 4.15.A**, left panel). Remarkably, these values exceeded the TRP of phage phylogenetic markers. For instance, for any phage-CLT pair, other phages of the same phylogenetic group had only an 18% chance of infecting the same bacterial capsular type (**Figure 4.15.A**, left panel). The accuracy of RBD-based CLT tropism prediction was variable, as 14 RBD clusters had a strong association with capsular tropism (TPR > 80%) whereas nine showed no association at all (< 10%) (**Figure 4.14.B**, right panel). We expected this considering that (1) many phages contained several Dpos, some of which could be cross-reactive against multiple capsular types, and (2) we could not experimentally test all the existing capsular types. Despite these limitations, we found 25 RBD clusters with a consensus capsular tropism, defined as clusters of RBD sequences in which at least two-thirds of the sequences belonged to phages spotting in the same CLT. These 25 clusters were composed of 79 RBDs of 65 phages and covered 19 distinct CLTs. RBD clusters involved phages from the same phylogenetic group (69.1%), but also from distinct genera and even families (22.7%; **Figure 4.15.B**), showing horizontal gene transfer of RBDs across large evolutionary scales.

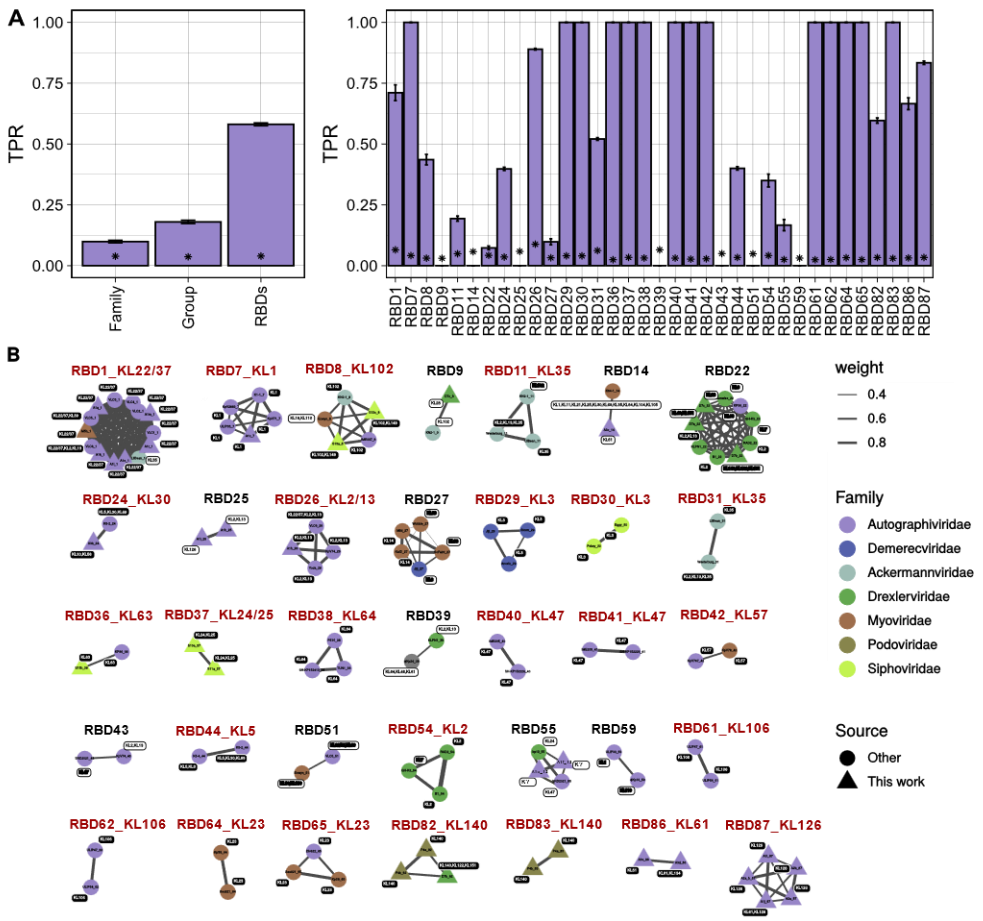

Figure 4.15. RBD-based prediction of phage capsular tropism. A. Left panel: Sensitivity or True Positive Rate (TPR) of different phage traits to predict infections in capsular types of bacteria. Bars represent the standard error of 3 independent calculations. Asterisks represent the mean TPR values obtained with randomized data (null expectation). For RBDs, the average of the 39 RBD clusters is shown. Right panel: TPR for each individual RBD cluster. **B.** Representation of the 39 RBD similarity clusters obtained. RBDs are named by the phage name followed by the phage protein in which they were detected. Clusters labeled in red are those for which a consensus capsular tropism could be obtained. Colors of nodes indicate phage taxonomic families. Sequences obtained in the present work are shown as triangles and previous published sequences as circles. Edge weight (width) represents amino acid identity between RBDs within the same cluster. Each RBD is annotated with the corresponding phage CLT tropism. White labels represent tropisms that did not match the consensus CLT of an RBD cluster.

4.2.4 Validation of sequence-based phage tropism using prophage sequences

To further analyze the predictive power of RBD clustering, we extracted a representative sequence from the 25 RBD clusters showing a consensus capsular tropism (**Figure 4.15.B**). We then interrogated the RefSeq database of bacterial genomes with RBD sequences in search of prophages with these motifs. We predicted the capsular tropism of the corresponding prophages by retrieving the CLT sequence of the host bacteria and inferring the capsular phenotype [159]. Then, we compared this prediction with the consensus CLT of the RBD cluster. We made enough predictions (>30) for 18 of the 24 RBD clusters (79%), which encompassed 15 distinct CLTs (KL2/13, KL3, KL22/37, KL23, KL25, KL30, KL35, KL47, KL57, KL63, KL64, KL102, KL106, KL126, and KL140). For the remaining clusters, we did not get enough hits with a confident CLT showing that these RBDs are rare in the prophages of typed bacteria.

Overall, we made ~10,000 predictions, distributed in 3,824 RefSeq bacterial genomes. The 18 RBD clusters had an average predictive power on CLT tropism of 35%, although this value varied widely, from 0% to 100% depending on the cluster. We found a statistically significant capsular tropism prediction for 13 clusters (72%). Predictabilities >98% were obtained for four RBDs-CLT associations (KL2/13, KL30, KL57, and KL63; **Figure 4.16**). Interestingly, it was possible to obtain accurate predictions (>99%) even beyond the *Klebsiella* genus (eg. *Escherichia/Shigella*, for CLTs KL2/13, KL30, KL35, KL47, KL57, and KL63; 4/4, 90/91, 6/6, 43/43, 80/81 and 23/23 correct predictions, respectively). Also, accurate predictions were found across a wide range of RBDs identities (**Figure 4.17**).

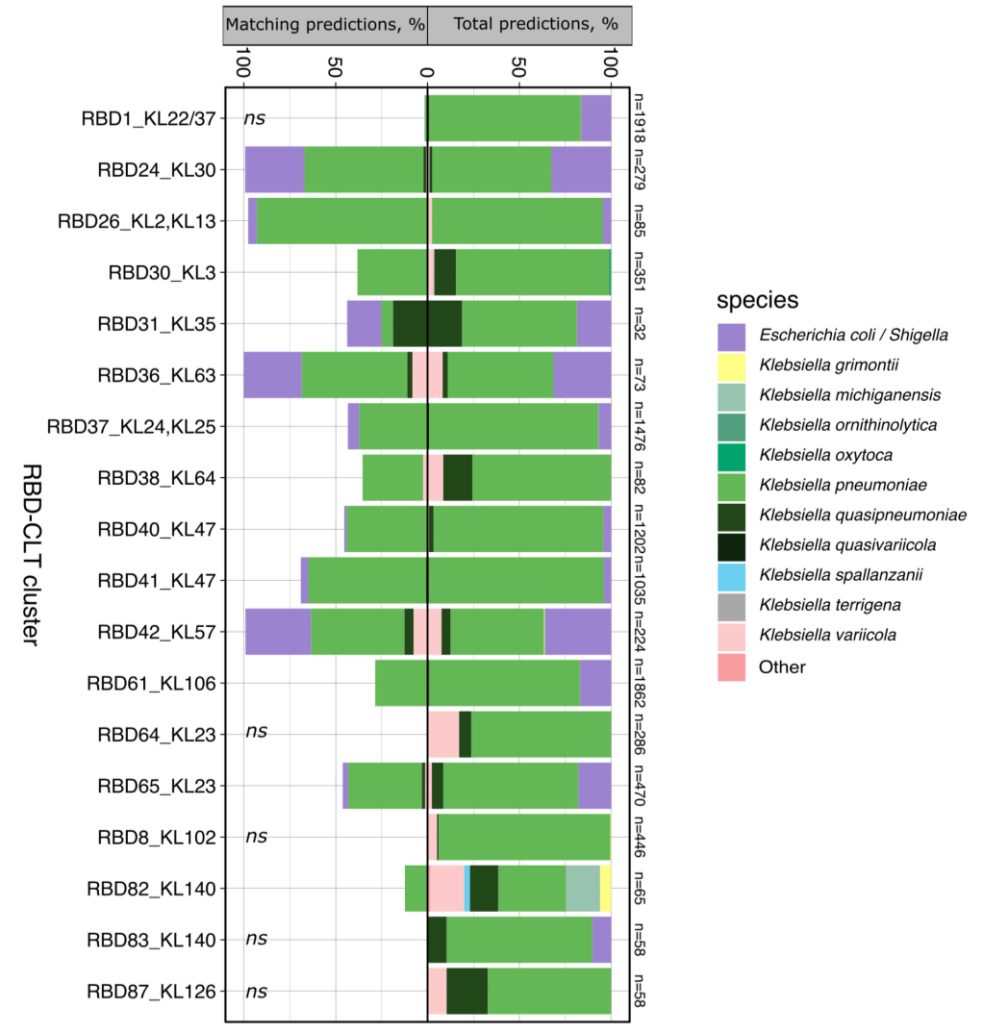

Figure 4.16. Validation of RBD specificity with prophage sequences. Ability of the consensus RBD clusters to predict the CLT tropism of prophages obtained from the RefSeq database as described in the main text. The percentage of total and correct predictions by bacterial species is shown. The n-values indicate the total predictions for each RBD. "Other" refers to spp. within *Klebsiella* sp. and also from other genera. *ns* denotes no significance by Fisher's exact test ($P > 0.05$).

For the five RBDs in which the predicted CLT based on prophages showed no significant association with the consensus CLT of the cluster analysis, we analyzed the distribution of hits to determine which factors might explain this failure. First, for RBD1_KL22/37, most hits (56.36%) were assigned to KL25 instead of to KL22/37. Interestingly, capsule recombination between both CLTs has been previously identified [149]. Second, for RBD8_KL102, we found that 55% of the predictions pointed to bacteria from the KL38. In this case, the RBD_KL102 representative sequence showed 35% amino acid identity and 84% query coverage with the depolymerase of phage ϕKp34, which infects KL38 [283], revealing a potential cross-reactivity or an evolutionary association between the abilities to infect both CLTs. Next, clusters RB64 and RBD65 were both present in phages infecting KL23, but only prophages with identity to RBD65 were found in bacteria from KL23, suggesting that the latter was the active depolymerase. The same situation was observed in the RBD83_KL140. Finally, for the RBD87 cluster, the predicted CLT based on prophages was KL14 (72%) instead of KL126 but the link between these two CLTs remains unknown.

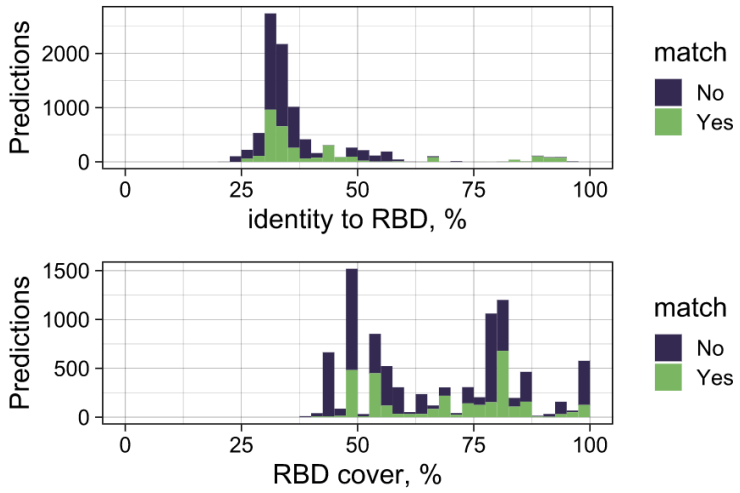

Figure 4.17. Tolerance of RBD sequences to predict phage capsular tropism. Distribution of RBD identity and coverage to consensus RBDs for prophage predictions. Match indicate if the CLT of the lysogenized host matched the consensus CLT of the RBD cluster used as a query. Hits with RBD ≤ 40 % RBD cover were not considered.

4.2.5. Prediction of productive infections

Spot tests provide a fast method for examining host tropism but they do not reveal productive infection [59,71]. To address this, for each of the 124 positive spot tests, we determined phage virulence by measuring optical density in liquid cultures. We detected a significant reduction in host density relative to non-inoculated controls in 94 of the 124 assays (76 %). Of the 30 negative pairs, five provided evidence of productive lytic infection by the progeny assay, and hence were also classified as virulent. Then, we reexamined host tropism by considering only the 99, out of 6319 (1.6 %), phage-host pairs yielding virulent infections in liquid culture (**Table S8**). Host breadth now ranged from 0 to 17 of the 138 tested strains. Again, the 42 non-S8/S9 phages were restricted to one or a few strains (median: 1, mean: 1.33), whereas S8-S9 phages exhibited a much broader range (median: 10.5, mean: 10.75). Host CLT was again the single factor that best predicted virulent infection, but with a considerably lower TPR than for spot tests (0.53 ± 0.01 for non-S8/S9 phages).

Therefore, spot tests seemed to be a better indicator of capsule-dependent host tropism than actual productive infection. To better understand this, we focused on the 25 phage-host combinations that yielded no-virulent infection despite the phages producing virulent infections in other strains (**Figure 4.18.A**). As expected, non-S8/S9 phages induced capsule digestion in the 13/13 phage-host combinations in which the phage produced a spot, despite being avirulent (t-test: $P \leq 0.06$; **Figure 4.18.B** and **Table S11**), showing the reliability of spot tests for assessing phage-induced depolymerization and capsule-dependent tropism. Interestingly, we observed successful adsorption for 16/25 combinations tested of which 14/19 included non-S8/S9 phages (**Figure 4.18.B**). Adsorption rates were close to 100% in many cases (**Table S12**), which would be difficult to explain with a lack of a *bona fide* receptor.

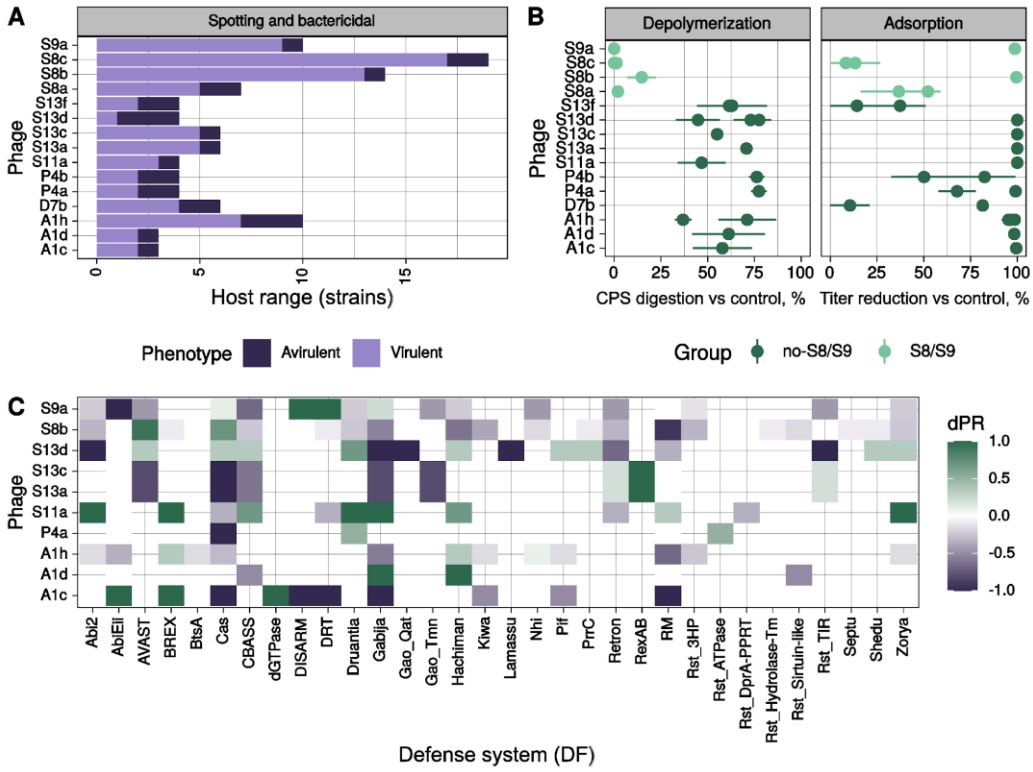

Figure 4.18. Study of avirulent infections. A. Summary of host-ranges for phages that showed both virulent and avirulent infections. Virulent infections were considered when the phage was able to produce a spot and to reduce bacterial density or produce progeny. Avirulence was considered when despite spot formation, no productive infection and no significant effect on host density were observed. **B**. Quantitation of depolymerase activity and adsorption in 25 avirulent phage-host pairs. Depolymerase activity was quantified by comparing capsule polysaccharide levels in phage-treated versus untreated controls. Adsorption was quantified by measuring phage titer following inoculation relative to cell-free mock cultures (eclipse phase). Each point represents the mean +/- SEM of at least 3 replicates. **C.** Differential probability of resistance (dPR) for each phage-defense system. For avirulent host-phage combinations, only those in which adsorption was observed (panel B) were considered. Values of 1 indicate that the defense system was exclusively found in avirulent combinations. On the contrary, negative values indicate that the defense system was overrepresented in virulent combinations and thus probably did not contribute to the observed resistant phenotype.

To investigate the causes of avirulence despite efficient phage adsorption, we searched genomes for known anti-phage systems. We found no CRISPR spacers or resident prophages with significant identity to the adsorbed phages, suggesting no involvement of these forms of immunity. To perform a more systematic analysis, for each phage, we calculated the frequency of known bacterial defense systems in the different strains. We found 39 phage-defense system combinations which were more frequent in resistant (phage avirulence) than in susceptible strains (dPR > 0). Of them, 14 were exclusive of resistant bacteria (dPR = 1; **Figure 4.18.C**). Overall, we found BREX, RexAB, Druantia, Hachiman, and dGTPase defense systems to be more frequently associated with resistant strains across phages (total dPR >= 1). Thus, even though non-S8/S9 phages stripped capsules and were adsorbed, intracellular mechanisms might contribute to explain the lower sensitivity of CLTs for predicting the productive host range of phages.

4.2.6. Determinants of infection of broad-range phages

Whereas, in our dataset, non-S8/S9 phages infecting a specific strain had a 92% probability of spotting other strains of the same CLT, this association dropped for the broad-range S8/S9 phages (TPR = 0.39 ± 0.01; **Figure 4.12.A**). Hence, the tropism of S8/S9 phages was weakly determined by CLT. As such, S8/S9 phages did not produce plaque haloes nor encoded canonical Dpos (**Table S7**). This suggests that S8/S9 phages use alternative mechanisms for penetrating the bacterial capsule. To verify this, we experimentally analyzed the depolymerase activity of 36 different host-phage pairs involving S8/S9 phages. In contrast to the strong capsular polysaccharide digestion activity shown by the other phages, S8 phages exhibited low depolymerization activity (**Figure 4.12.B**). However, we found some capsule-degrading activity for some strains. For instance, phage S8b depolymerized strain KL39-CU293 but did not show depolymerase activity in KL39-CU616 (**Table S9**). Interestingly, we found the opposite pattern in phage S8c, even though neither phage produced haloes in any of the tested hosts (**Table S7**). Thus, some domains found in the genomes of S8 phages might be active against carbohydrates.

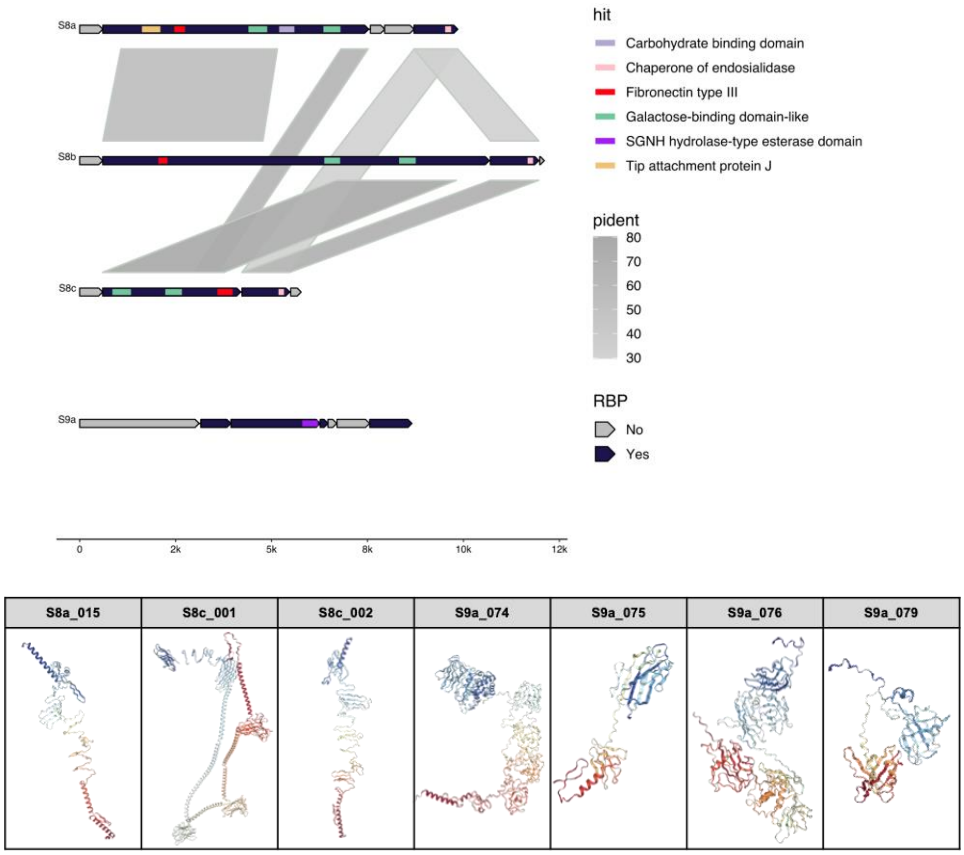

Figure 4.19. RBPs of S8/S9 phages. Top: Genomic comparison of S8/S9 RBPs. Domains found with InterProScan5 are shown. Pident shows the aa percentage identity after blastp comparison. Bottom: Predicted RBPs secondary structures. Only structures of RBPs < 1500 aa are shown. Each structure is labelled with the phage name and the protein id. S8a_015 corresponds to the second RBP of S8 phages.

To explore the S8/S9 encoded mechanisms of host recognition, we analyzed their putative RBPs. We found that S8 phages encoded two RBPs each, while the S9a phage had four (Table S13). Interestingly, the S8a and S8b first RBPs had > 1,500 aa (2,309 aa and 3,201 aa, respectively), suggesting that they might represent a polyprotein [260,284]. The first RBP of S8 phages showed 2-3 carbohydrate-binding modules (CBMs), whilst the second RBP showed a chaperone domain of endosialidase (Figure 4.19 and Table S13). However, no catalytic domains were found, nor the typical structure of depolymerases with parallel β-sheets when modelled with AlphaFold2 (Figure 4.19). Differences in

RBPs domains might be responsible for the partial overlapping tropism of S8 phages (**Figure 4.11** and **Table S8**). In the case of the phage S9a, only two RBPs showed > 500 aa, suggesting that the other two might not have host recognition activity. For the remaining two RBPs, one showed a SGNH hydrolase-type esterase domain, suggestive of endolysin activity rather than Dpo activity [285].

Contrarily to non-S8/S9 phages, the ability of the CLT to predict productive infection increased for broad-range phages (spotting TPR: 0.39; productive infection TPR: 0.46). In total, there were six S8 phage-bacteria combinations that showed a positive spot but that did not yield a productive infection in liquid (**Figure 4.18.A**). These were not due to depolymerization activity nor successful adsorption of S8/S9 phages (**Figure 4.18.B**). These spots were very turbid, which may suggest phage replication in a subset of bacterial cells [286]. We thus hypothesized that S8/S9 turbid spots might be the result of phage replication in a subset of acapsular or low-capsular-producing bacteria [169]. To examine this possibility, we compared phage spots in wild-type encapsulated bacteria, a spontaneous acapsular mutant (Cap-) and in a 1:1 mixture of both. For 4/6 combinations, we observed that S8 phage spots became clearer when increasing the concentration of the Cap- mutant in the agar plate (**Figure 4.20**). This suggests that, for some hosts, S8/S9 phages use alternative modes of entry that remain uncharacterized but are likely independent of capsular carbohydrate binding.

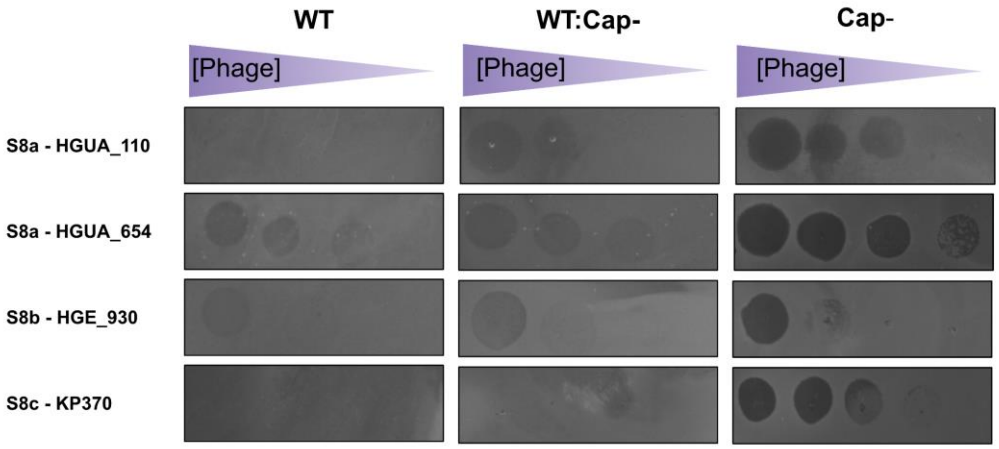

Figure 4.20. Non-capsular determinants of host-tropism of S8/S9 phages. Spotting of serial dilutions of S8/S9 phages (10-10^4 PFUs) as a function of the fraction of acapsular bacteria (Cap-) in the plate. Three independent replicates were performed and a representative image was chosen. WT: WT-bacteria only. WT:Cap-: mix of WT-Cap- (1:1). Cap-: Cap- only. Combinations CU630-S8c and NTUH-S9a were omitted as spots were not observed with these concentrations.

4.3 Beyond specificity: the role of capsules for *Klebsiella*-phage interactions

In the previous section, we found two main patterns of phage specificity: **capsular type-specialists** (non-S8/S9, n=42) and **capsular-independent generalists** (S8/S9, n=4). To simplify, we will hereafter refer to capsular-type specialists as **narrow-range** phages and to capsular-independent generalists as **broad-range** phages. Narrow-range phages have tropism toward specific types of capsules, suggesting that capsule sugars act as receptors. On the other hand, we have shown that broad-range phages could benefit from the removal of this layer due to alternative mechanisms of action. The extent to which the capsule is essential for both types of phages is unknown. In this last section, we will address this question systematically by testing phage infection against acapsular bacteria. These acapsular bacteria have been derived from susceptible hosts to determine whether capsule loss is enough to prevent phage infection. In addition, we will explore whether the role of the capsule for phages has consequences on bacteria and phage growth dynamics.

4.3.1 Capsular heterogeneity and phage host-range

Phage infectivity was previously determined using wild-type, encapsulated bacteria (WT-Cap+), which occasionally undergo capsule inactivation during cultivation. A total of 75 strains were productively infected by at least one phage and, therefore, were selected for this study (**Figure 4.21** and **Table S14**). To determine the extent of capsule inactivation, overnight cultures of the 75 WT-Cap+ strains were diluted and plated to obtain single colonies. While the appearance of WT-Cap+ colonies is mucoid and opaque, some translucent colonies or sectors can appear, indicative of capsule-disrupting mutations [287]. We found the frequency of Cap- mutants ranged from 0 to 100 (mean: 8.97, median: 1.12) per 100 CFU of WT-Cap+ (**Figure 4.22.A** and **Table S14**). Of the 75 bacterial strains, 3 were actually non-capsulated (100 Cap- per 100 CFU), and for 7 strains (9%) we were unable to obtain any Cap- mutant (**Figure 4.21** and **Table S14**). Therefore, 90% of

the *Klebsiella* cultures presented phenotypic heterogeneity with the coexistence of encapsulated and non-encapsulated clones.

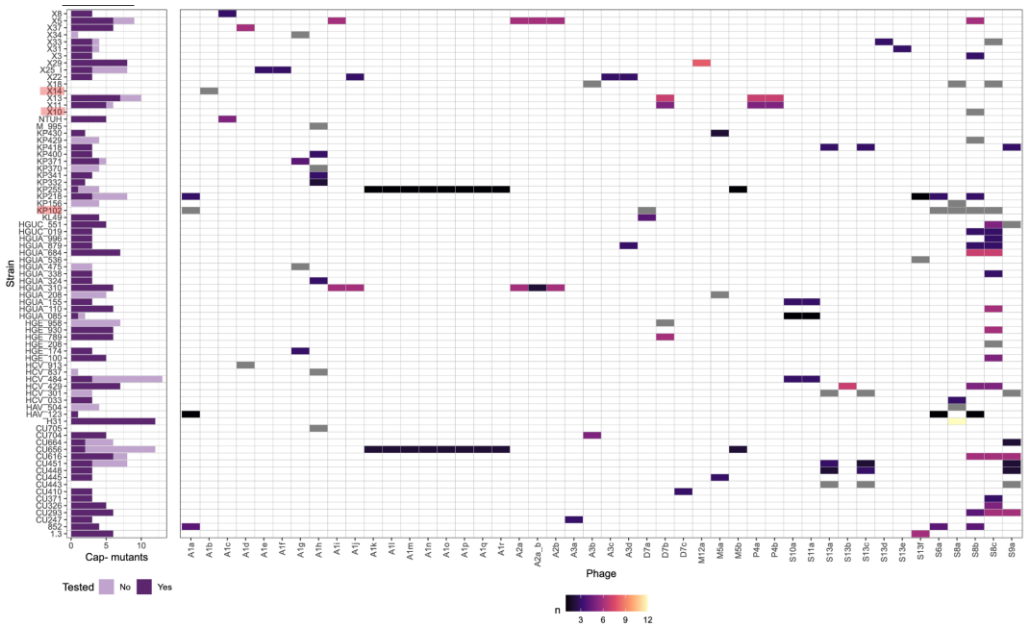

Figure 4.21. Overview of the Cap- collection. Each row represents a WT bacterial strain and each column a phage. Only bacteria infected by at least one phage are shown (n=75). Red labels indicate that the morphology of the WT bacteria was acapsular. Left panel: number of stable Cap- mutants isolated for each bacterial strain. Some mutants were not tested. Right panel: Number of Cap- mutants tested for each possible combination of bacteria and phage. Combinations of phage-bacteria in which Cap- mutants were not tested are colored in grey. Broad-range phages are those of groups S8/S9 (S8a-c and S9a).

To determine if capsular heterogeneity affected phage tropism, we compared the proportion of Cap- mutants in bacteria infected by broad- or narrow-phages. We found that bacteria susceptible to broad-range phages had ~ 20 times more Cap- mutants than resistant bacteria (6.95 Cap-/100 CFU *vs.* 0.87 Cap-/100 CFU, *P*=0.02). Narrow-range phages showed the opposite trend, with 7.98% and 12.97% of Cap- mutants in susceptible and resistant bacteria, respectively (*P*=0.56) (Figure 4.22.A). This contrasting preference could lead to a small overlap in infections by both types of phages. Only 12 bacterial strains were infected by

both narrow- and broad-range phages (Figure 4.22.B). Concretely, the P(Infection$_{narrow}$ | Infection$_{broad}$) was 14.4%, while the P(Infection$_{broad}$| Infection$_{narrow}$) was even lower, 7.59%. Hence, the host-range of broad-range phages was biased towards bacterial populations with acapsular bacteria.

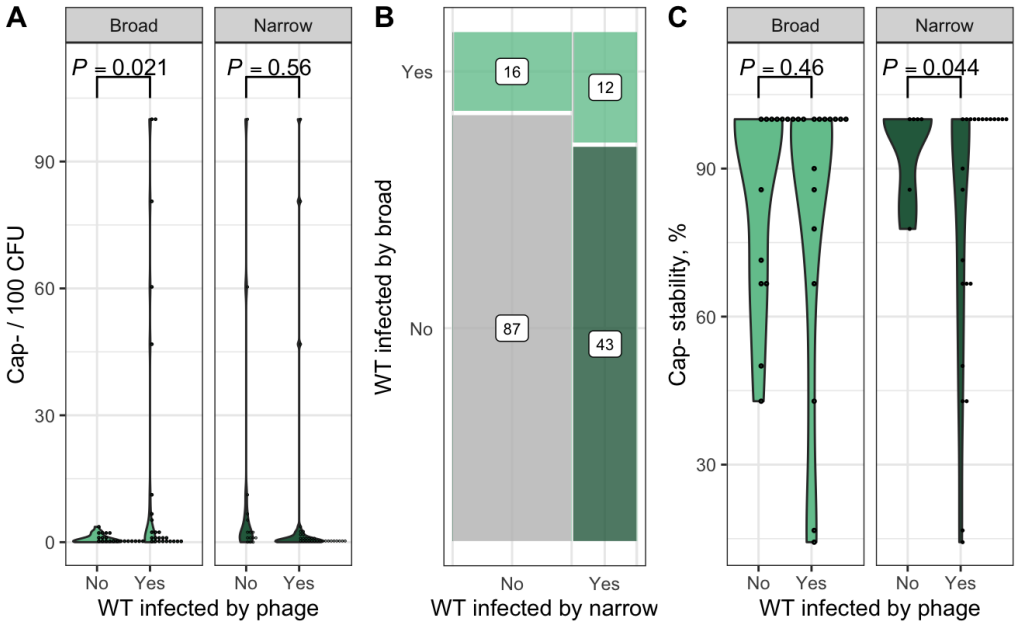

Figure 4.22. Capsule heterogeneity and patterns of infection. A. Spontaneous Cap- colonies per 100 CFU, depending on bacterial infection status by broad- or narrow-range phages. Each point represents the mean of at least 2 independent replicates for each bacterial strain (n=45). **B.** WT bacterial strains infected by narrow-range, broad-range, or both types of phages in WT bacteria. The total number of cases for each combination is shown. **C**. Stability of Cap- mutants. Each point represents the mean for each bacterial strain having at least three Cap- mutants with the stability recorded (n=31). P-values were obtained with the t.test ().

To confirm that the acapsular phenotype was stable, Cap- mutants were restreaked and their morphology recorded. Cap- stability was quantified as the fraction of acapsular bacteria that was picked and remained non-capsulated after one passage for each WT-Cap+ strain. We found that Cap- mutants were mostly stable (mean: 0.83, median: 1), confirming a genetic basis for capsule inactivation. Cap- stability did not differ between bacteria susceptible to narrow- (Cap- stability

= 0.79) or broad-range phages (Cap- stability = 0.78) phages. However, the stability of Cap- mutants of bacteria resistant to narrow-range phages was 19% higher (Cap- stability=0.94)(Figure 4.22.C). This result suggests that narrow-range phages preferentially infect bacteria in which the loss of the capsule is not stably inherited.

4.3.2 The basis of Cap- mutants infectivity

Next, to determine how capsule inactivation affects phage resistance, stable Cap- mutants were challenged to phages by the bactericidal assay (§3.6.2). We tested 223 Cap- mutants derived from 55 different bacterial strains, covering 112 out of the 137 (82%) possible combinations of WT-Cap+ and phage (Figure 4.21). A Cap- mutant was scored as resistant if the phage-treated bacteria grew similarly to the untreated Cap- mutant. As a control for phage infectivity, the original WT-Cap+ bacteria were also challenged in parallel. We found that Cap- mutants acquired resistance to phages in 286 of the 406 combinations tested (70%) (Figure 4.23 and Table S15), confirming that the capsule is not only a determinant but it is also required for successful infection by *Klebsiella* phages. Surprisingly, half of the Cap- mutants challenged with broad-range phages were resistant, while the corresponding encapsulated bacteria were not (Figure 4.23). These results further suggest the use of both capsule and non-capsule receptors by these phages.

Contrarily, we found 120 combinations in which Cap- mutants remained susceptible to infection (30%), revealing the use of capsule-independent receptors. Broad-range phages represented the majority of Cap- mutants infections, with 83 cases, but narrow-range phages were also able to infect non-encapsulated bacteria (Figure 4.23). Phages capable of using capsule-independent receptors were distributed across groups, showing that this is a character with low phylogenetic association (Figure 4.23). S9a for broad-range phages, and D7a and S13f, for narrow-range phages, showed the opposite trend than expected based on their host-range category. That is, phages D7a and S13f infected all Cap- mutants tested (4 and 7, respectively), while phage S9a infected none (n=21). Additionally, there were 12 phages (A1a, A1d, A1g, A1i, D7b, M12a, S10a, S13e, S6a, and S8a-S8c) which only infected a subset of Cap- mutants (Figure 4.23).

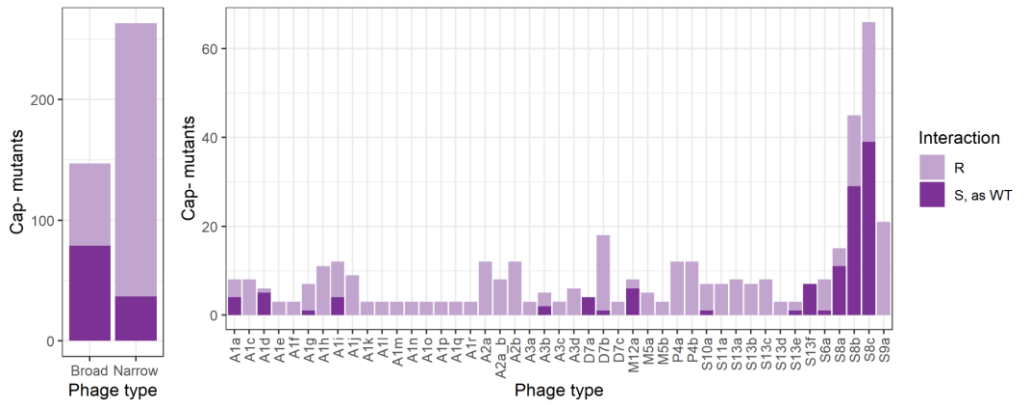

Figure 4.23. **Infectivity of Cap- mutants.** Left: Susceptibility of Cap- mutants to broad- (S8a-c and S9a) and narrow-range phages. Right: Susceptibility of Cap- mutants decomposed by individual phages. R: Resistant, S: Susceptible. Phage A1b was not included because it only infected WT non-encapsulated bacteria. A total of 410 Cap- mutants x phage combinations were tested with a minimum of 3 Cap- mutants per phage.

To better understand the above exceptional cases, we further examined the RBPs of these phages [260] in search for non-Dpo domains that could recognize other bacterial receptors apart from the capsule. The predicted structures of RBPs were compared with those from databases and the TM-scores obtained [265](§3.5.5). TM-scores indicate the similarity of two structures, with values ≤ 0.2 corresponding to unrelated proteins and scores ≥ 0.5 to structures with the same fold [288]. The findings for each phage are detailed next.

First, phage S9a encodes 4 RBPs, 2 of them predicted to be functional due to their size (Figure 4.19). The first RBP (#S9a_074) was similar to a GDSL-like lipase/acylhydrolase (TM = 0.44, Figure 4.24), which can possess lipase and esterase activities [289]. Esterases can deacetylate surface polysaccharides, being essential for phage infection but leaving the backbone of the polysaccharide intact [174]. This would explain why phage S9a could not infect any Cap- mutants and the lack of depolymerization activity (Table S9). The second RBP (#S9a_76) was related to the host specificity protein of *Salmonella* prophage Gifsy-2 (TM = 0.38, Figure 4.24), which uses the protein OmpC as a receptor [290].

Second, phage D7a also encodes for a RBP protein with an esterase domain (#D7a_045) (**Figure 4.24** and **Table S16**) but it might use its additional RBP (#D7a_043), similar to the tail protein of prophage Gifsy-2 (TM = 0.57, **Figure 4.24**), to infect acapsular bacteria. This RBP showed similarity to a leucine-fold (TM = 0.54), further supporting its role in mediating protein-protein interactions [291]. The closely related phage D7b, also capable of infecting Cap- mutants, shares this second RBP (**Figure 4.24**).

Third, phage S13f was the only phage not adsorbed despite successful capsule depolymerization (**Figure 4.18.B**). This already suggested that the capsule was not a receptor for this phage, in agreement with the results found with the Cap- mutant infectivity assay. The tail protein of this phage was also similar to that of Gifsy-2 (TM = 0.59), suggesting recognition of proteins. Additionally, this RBP had a similar fold to an endo-1,4-beta-xylanase 5 (**Figure 4.24** and **Table S16**), which might explain the depolymerization activity of this phage lacking a canonical Dpo (**Table S7**).

Finally, we found that phages M12a and S10, also capable of infecting Cap- mutants, encoded folds likely involved in protein-protein interactions due to identity to lectin-domains and the Gifsy-2 host specificity protein (**Figure 4.24**). For phages A1a, A1d, A1g, and A1i we could not detect additional domains in RBPs other than Dpos. This could indicate that Cap- mutants infected by these phages have not completely lost their capsule. Alternatively, the phages might bind to the LPS using carbohydrate-binding-domains of Dpos [292].

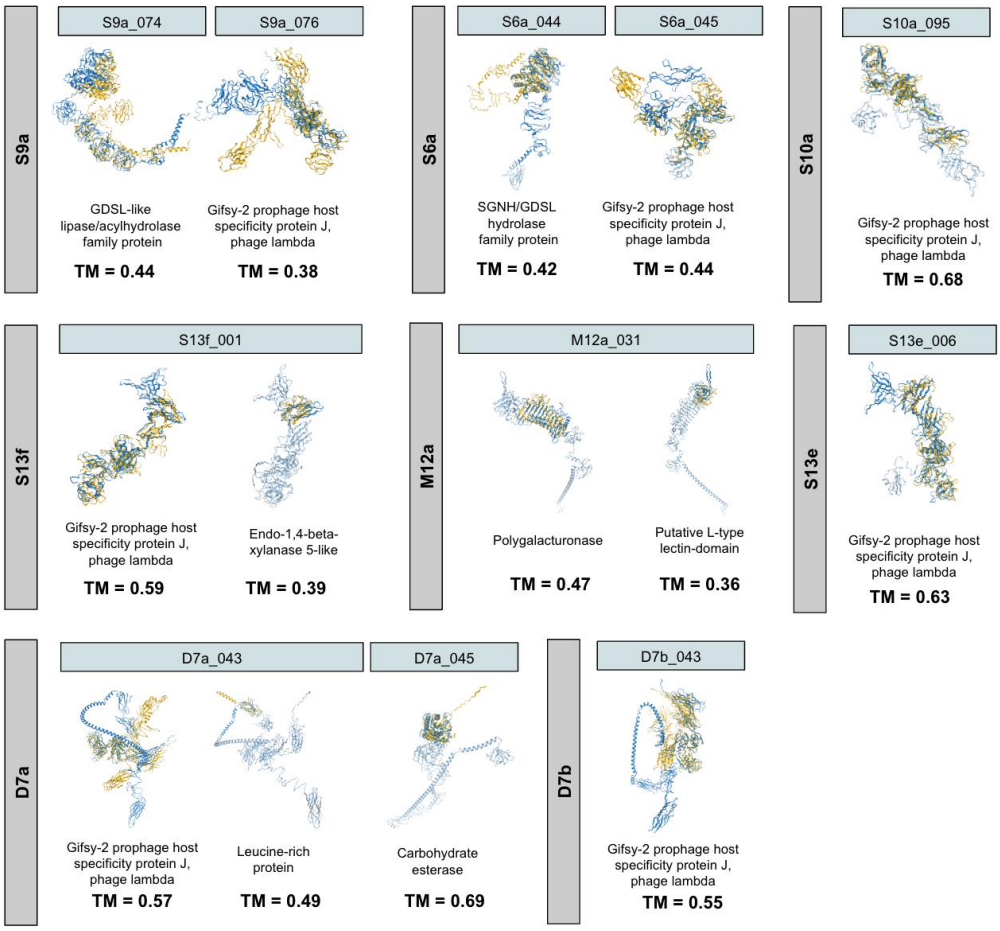

Figure 4.24. Structural alignments of phage RBPs without Dpo activity. Structures of RBPs of phages of this work are shown in blue. Structures from the database (afdb-proteome) are shown in yellow. Blue boxes indicate the phage protein identifier. TM values indicate the TM-score between pairs of structures.

In summary, the capsular tropism and RBPs of phages only partially explained the pattern of infection towards Cap- mutants.

4.3.3 Contingency of Cap- infectivity

Apart from phage factors, the infectivity of Cap- mutants may be affected by the genetic background of the bacteria. We hypothesized that phages might use the capsule or alternative receptors based on the CLT. For example, some phages

might recognize certain CLTs by using Dpos (Figure 4.15) or CBM domains (Figure 4.19), while for others CLTs they may use different RBP domains (Figure 4.24), therefore not requiring the capsule. To test this hypothesis, we assessed the probability that phages exhibit the same infection phenotype (pEqual) depending on whether Cap- mutants were grouped by CLT, OLT, ST, or bacterial strain (i.e., the original WT-Cap+ from which the mutant was obtained). We only considered phages that infected at least one Cap- mutant (n = 13) to remove biases of phages lacking RBDs apart from Dpos. We found no significant association between pEqual and the grouping factor (ANOVA, $P = 0.68$). However, bacterial strain was slightly a more important factor shaping Cap- mutants' infectivity within a phage (pEqual-Strain: mean = 0.51; median = 0.56) than the CLT (pEqual-CLT: mean = 0.43; median = 0.43) (Figure 4.25).

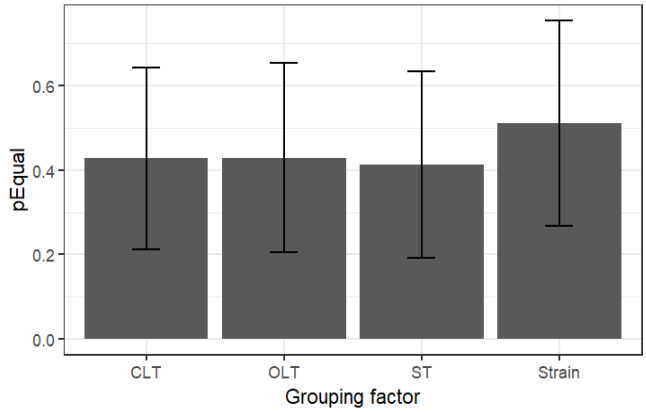

Figure 4.25. Probability of the same Cap- infection outcome (pEqual) based on bacterial traits. The height of the bars indicate the mean value and error bars the standard deviation between phages (n=13). pEqual was determined from all possible pairwise comparisons of Cap- mutants' within a phage, dividing the number of comparisons with an equal outcome (infection or resistance) between the total number of comparisons. CLT: Capsular Locus Type. OLT: O-antigen Locus Type. ST: Sequence Type.

Phage A1a, for example, was able to infect all Cap- mutants of strain KP218 (n=3) while all Cap- mutants of strain 852 (n=4) were resistant to the phage, both of which belonged to CLT KL151. Another example is provided by phages S8b and S8c: phage S8b was not able to infect any Cap- mutant of the strain KL39-CU293 (n=6), indicating a capsule-dependent mode of entry. In contrast, this same phage infected all Cap- strains of another strain from the same capsular type KL39 (CU616, n=6), indicating capsule-independent tropism for this strain. The opposite pattern was found for the related phage S8c tested against the same subset of Cap- mutants.

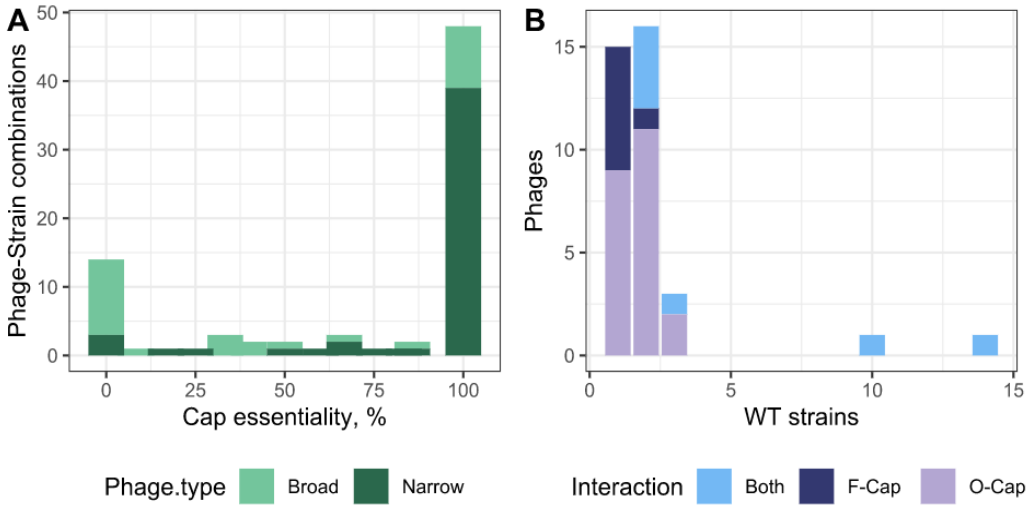

Figure 4.26. Contingency of capsule essentiality for phages. A. The capsule essentiality value per each phage and WT-Cap+ bacterial strain is shown. Capsule essentiality represents the number of resistant Cap- mutants between the total of Cap- mutants tested. Only bacteria for which at least 3 Cap- mutants were tested are shown. **B.** Classification of phages based on the number of susceptible strains. Both indicate that the phage showed facultative capsular (F-Cap) and obligate capsular (O-Cap) interactions.

Overall, 7 out of 21 (33%) phages infecting more than one bacterial strain showed both capsular-obligate (none Cap- mutant infected from a WT-Cap+ strain) and capsular-facultative tropism (at least one Cap- mutant infected) (Figure 4.26). This might indicate that different mechanisms of capsule inactivation lead to distinct infection phenotypes as mechanisms are strain-specific [147]. Additionally, we found 18 cases in which phages could infect only some, but not all, Cap-

mutants from the same WT-Cap+ strain (Figure 4.26). This further supports that different mechanisms of capsule inactivation may have distinct effects on phage infection.

4.3.4 Quantification of phage-bacteria dynamics

Finally, to study how capsule-mediated interactions impact phage dynamics, we analyzed WT-Cap+ bacteria growth curves in the presence of phages. Bacterial killing curves are usually characterized by a first period of phage lysis, producing bacterial growth arrest (Figure 4.27.A). Alternatively, phages with lower virulence can take longer to control the growth of the bacterial population, and the lysis effect of these phages is delayed (Figure 4.27.B). This lysis period is usually followed by bacterial regrowth, resulting from a substantial proportion of the bacterial population being able to evade phage predation (Figure 4.27A-B). In some cases, bacterial regrowth is not observed during the time course analyzed (Figure 4.27.C), indicating that phages are able to suppress resistance. By leveraging on these distinctive dynamics, bacterial killing curves have been extensively used to inform about phage-bacteria dynamics and resistance emergence [293–295].

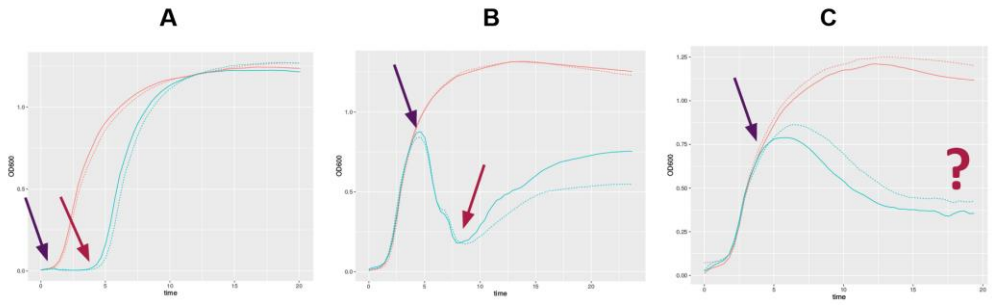

Figure 4.27. Examples of bacterial growth curves based on optical density measurements (OD600) in the absence (red) or presence (blue) of phage. Arrows denote the point of resistant regrowth. **A.** High virulent phage with bacterial regrowth. **B**. Low virulent phage with bacterial regrowth. **C.** Low virulent phage with no detectable bacterial regrowth. The purple arrow shows the time of phage lyis (t_{lys}) and the red arrow denotes the time of bacterial regrowth (t_{res}).

Comparison of the area under the curve (AUC) of phage-treated and untreated bacteria can be used to obtain the percentage of inhibition of phages (pIAUC) [269]. Higher pIAUC values indicate greater phage control of bacterial growth over the entire time course and lower pIAUC are indicative of less efficient phages. Overall, we found that phages were inefficient at suppressing *K. pneumoniae* growth *in vitro*, with a mean pIAUC of 35% (**Table 4.1**). To correlate this with the role of the capsule in a given phage-bacteria interaction, we determined the capsule essentiality value for each pair. Capsule essentiality was defined as the ratio of resistant Cap- mutants among the total of Cap- mutants tested. Therefore, values of 1 indicate that the phage did not infect acapsular bacteria of that particular WT-Cap+ (**Figure 4.26.A**). We found a negative correlation between capsule essentiality and pIAUC (Spearman R=-0.36, $P < 0.01$)(**Table 4.1**), suggesting that when phages use the capsule for infection they are less efficient.

Table 4.1. Correlation of phage-bacteria dynamics parameters with the capsule essentiality. **: P < 0.01, *: P < 0.05. V: Virulence. R: Resistance.

Parameter	mean	sd	Spearman R	P-value	Informs of
pIAUC (%)	35.01	17.57	-0.364	0.001**	V, R
t_{lys} (hpi)	2.09	0.67	-0.110	0.340	V
t_{res} (hpi)	6.48	4.00	-0.230	0.041*	R
t_{eff} (hpi)	3.58	2.66	-0.287	0.015*	V, R
pIAUC-VIR (%)	81.58	20.72	0.097	0.3959	V
pIAUC-RES(%)	35.39	29.01	-0.309	0.006**	R
Plaque size (mm)	1.04	1.08	-0.11	0.544	V

To quantify the phage-bacteria dynamics more precisely, we determined the time of observable bacterial lysis (t_{lys}), and bacterial regrowth after phage lysis (t_{res}). We obtained these by calculating the growth difference (ΔOD_{600}) between the phage-treated bacteria and the untreated control. Negative values indicate bacterial growth impairment due to the phage, and therefore, lysis. If resistant bacteria emerge, growth differences (ΔOD_{60}) between phage-treated and untreated controls start to diminish (Figure 4.28.A), being ΔOD_{600} values less negative. Therefore, we extracted the first negative point as a proxy of t_{lys} and the first negative valley as a proxy of t_{res} (§3.7.4). The time from phage lysis (t_{lys}) to resistance emergence (t_{res}) was denoted as the duration of effective control of phages (t_{eff}) (Figure 4.27). We found a good fit in most phage-bacteria combinations, both for t_{lys} and t_{res} (Figure 4.28), suggesting a broad application of our method to model phage-bacteria dynamics. Next, based on the inflexion point of resistance, we divided the pIAUC of phages in two phases: first, previous to the emergence of resistance, and therefore informative about phage virulence (pIAUC-VIR), and second, after the emergence of resistance to the end, and therefore more informative about the cost of phage resistance (pIAUC-RES)(Figure 4.27).

Overall, phages started to lyse bacteria fast (t_{lys} ~ 2 hpi) but resistant bacteria started to emerge after only ~ 4 hours (Table 4.1). We found a negative correlation ($P < 0.05$) of capsule essentiality with dynamic parameters indicative of bacterial resistance (t_{res} and pIAUC-RES) but not with parameters indicative of phage virulence (t_{lys} and pIAUC-RES) (Table 4.1). Therefore, capsule-mediated interactions do not affect the initial stages of phage lysis but, instead, phages seem to be constrained by a rapid emergence of resistance. To further corroborate this, we analyzed the size of phage plaques, as they represent a measure of phage virulence [296]. We did not find a significant correlation of phage plaque size and capsule essentiality ($P = 0.544$). To sum up, our results suggest that capsule-obligate interactions limit the co-evolution of phages possible due to a high frequency of capsule inactivation, and therefore, phage resistance. Further experiments will be required to ascertain this.

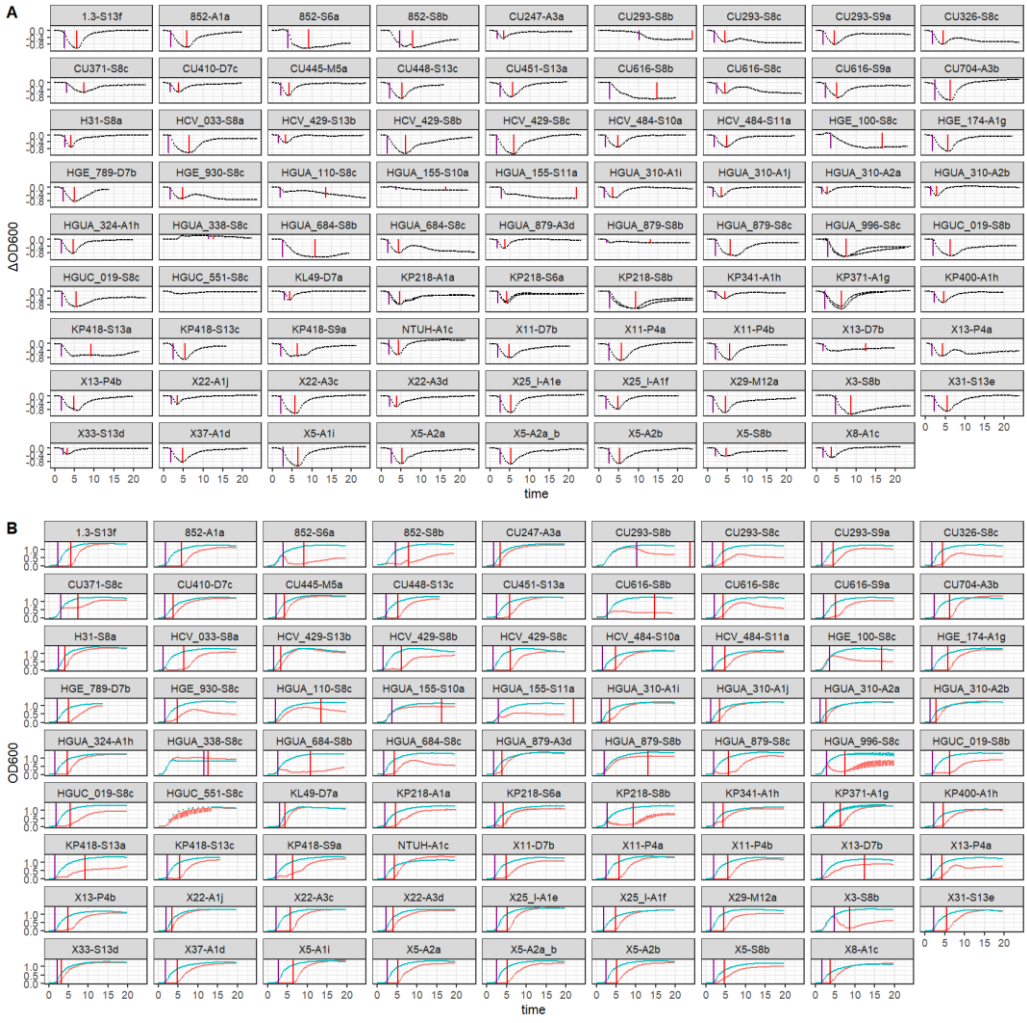

Figure 4.28. Quantification of phage-bacteria dynamics in WT-Cap+ bacteria. A. ΔOD_{600} (OD_{600}infected - OD_{600}non-infected). **B.** OD_{600} for infected (red) or non-infected (blue) bacteria. For visualization purposes, the average of ≥ 2 replicates is shown. The purple vertical line indicates t_{lys} and the red vertical line indicates t_{res}.

5. Discussion

5.1 The need for bacteria and phage collections

The lack of representative phage and host collections is one of the main obstacles to study phage-bacteria interactions [72,142]. Phages that infect *E. coli*, such as *Escherichia* "T phages" (from T1 to T7) [297] and bacteriophage lambda [298], are among the most studied due to their use as models in molecular biology. Due to the increasing awareness of phage diversity, phage research has now shifted to non-traditional phages. This shift has also been driven by the need for alternative therapies, such as those based in lytic phages, to treat antibiotic-resistant infections [12,28]. More than 20,000 phage genomes are currently available [39]. However, most of them are unpublished or poorly characterized, which prevents gaining further insights into the mechanisms of phage-bacteria interactions [142]. Importantly, many phages have been isolated using a common bacterial strain. For example, hundreds of phages have been isolated using *E. coli* K-12 [82,299], and thousands of them with *M. smegmatis* mc2155 [82] under the SEA-PHAGES program. Phages infecting a common bacterium can be highly diverse [45,82], but the strong bias towards some bacteria needs to be addressed.

Another constraint to study the determinants of phage tropism is the lack of host genome sequences [142]. This can result in phages being incorrectly classified as having a broad host-range when they may only infect a narrow range of closely related bacteria (for instance, derived from clonal expansions in hospital settings or environmental blooms). As such, it is important to consider the genetic diversity of hosts. Recently, several phage-bacteria collections with phenotypic and genomic data have been published. Kauffman *et al.* described the infection matrix of 248 *Vibrio* phages in 279 hosts that were isolated from seawater [54,300]. Likewise, Piel *et al*. studied the determinants of phage host-range using 195 strains of *V. crassostreae* and 243 vibriophages [113]. For Gram-positive bacteria, Göller and colleagues isolated 94 staphylococcal phages from wastewater and tested their host range in a diverse collection of 117 staphylococci [301]. Despite this trend, diverse phage-bacterial collections with genomic data are still needed for most hosts.

K. pneumoniae is a Gram-negative bacterium found in various environmental niches, such as water, soil, and plants. It is also commonly found in

the human gut microbiome as a commensal bacterium. *K. pneumoniae* can cause severe infections in immunocompromised patients, being an important nosocomial pathogen [191]. Compared to other Gram-negative opportunists, *K. pneumoniae* presents a higher genomic diversity [202]. One of the main hotspots of genetic diversity in *K. pneumoniae* is the capsular locus, with 77 serotypes and more than 180 CLTs described so far [156]. This diversity results from point mutations, genomic rearrangements, and extensive recombination [159]. Additionally, *K. pneumoniae* can undergo capsule inactivation [149,287], which can be beneficial for the bacteria, allowing them to thrive in some environments such as in epithelial cells or the urinary tract [154]. For all these reasons, *K. pneumoniae* has become a model to study the fitness effect of capsules across environments [147,149,302]. The bacterial capsule is the first barrier encountered by many phages. Its role in phage infection can range from offering protection [177] to acting as a receptor, allowing infection [189]. As such, *K. pneumoniae* also represents a great model to study the co-evolution between bacterial capsules and phages.

To investigate the effect of capsules on phage infection, we first sought to establish a representative collection of *K. pneumoniae* strains that spanned a wide range of capsular diversity. To achieve this, we used genomic surveillance data on antibiotic resistance in the Comunidad Valenciana region [208]. We expected 100 different CLTs by analyzing ~ 1,000 genomic sequences [159] but we found 73 CLTs (**§4.1.1**). This number was still high given that we only focused on clinical strains from a specific geographical region. Nevertheless, including environmental samples and/or samples from other regions [303] would have likely increased the capsular diversity. Despite this, our collection represents the diversity of the species (**§4.1.1**), making it a valuable resource for studying phage host-range and beyond that.

Given the ubiquitous distribution of *K. pneumoniae*, phages against this bacterium can be isolated from environmental samples such as sewage [204]. The EnBiVir lab (I2SysBio, UV-CSIC) provided 71 phage plaques that were isolated using *K. pneumoniae* hosts from soil and water samples. We successfully sequenced and characterized 70 of them. This number represents an increase of 13% in the number of *Klebsiella* phages available to date (n=555, accessed on July 22, 2022). Overall, this study has the largest number of *Klebsiella* phages described to date, followed by Lourenço and colleagues´ recent preprint with 68

Klebsiella phages [207], the work by Townsend *et al*., in which they reported 30 distinct *Klebsiella* phages [206], and the work of Beriot *et al.,* with 21 [304]. After genomic analyses, our phages were classified into 46 strains, 45 species, and 13 similarity groups (A1-A3, P4, M5, S6, P7, S8-S11, M12, and S13). All phage collection sites were very close, which could explain the presence of similar phages at different sampling points. Nevertheless, the presence of both highly related and unrelated phages from samples also reflects a high level of coexistence of phages [299,305].

Specifically, the phage collection includes viruses of the 5 main families of *Klebsiella* phages (*Autographiviridae, Drexlerviridae, Myoviridae, Podoviridae,* and *Siphoviridae)* (§4.1.5). Of the 45 phage species found, only one was previously deposited in databases, with no associated publication (https://www.ncbi.nlm.nih.gov/nuccore/940326120). Hence, this work contributes 44 new phage species. In addition, we report 11 phages with genomic similarity values to database sequences below 70%, therefore representing new genera [21]. Interestingly, this discovery rate was higher than in similar studies. For example, Townsend *et al.*, found that 15 out of 30 phages had ANI values ≥ 95% with the database sequences. Also, only 3 phages represented new genera [206]. The isolation strains, which are crucial for phage discovery, may explain the differences between similar studies. The phages of this work were isolated using 36 different clinical strains of *Klebsiella* (§4.1.3), as opposed to 24 hosts of the aforementioned work. This is likely related to the diversity of bacteria included but also to the permissiveness of the hosts for phage infection. The lytic cycle of the phage must be complete in order to isolate phages from plaques, a requirement that is only met for a limited number of bacteria-phage pairs. As a result, the use of bacterial strains with a restricted repertoire of intracellular phage defenses could help in the isolation process. A good example of this approach is the one developed by Maffei *et al*., in which they isolated phages against *E. coli* K-12 deprived of RM systems [82]. Regardless of the isolation method, our findings show the importance of carrying out more extensive sampling to discover the diversity of phages and, in particular, those infecting *Klebsiella*.

5.2 Systematic analysis of *Klebsiella*-phage interactions

The host range of a phage is highly dependent on the bacterial strains used for its evaluation. Because we cannot test all possible bacteria, it is crucial to select a host panel that is biologically relevant to the hypotheses being tested [72]. In previous studies with *Klebsiella* phages, hosts were restricted to the reference collection of *Klebsiella* serotypes [59,171], clones of multidrug-resistance [193], or to some species of the *K. pneumoniae* complex [173,206]. These studies, although highly significant, led to host-range determination in a few capsular types and/or a single strain within a capsular type. To overcome this, we studied phage tropism in the aforementioned *K. pneumoniae* collection, composed of 138 representative strains from 59 CLTs, 33 of which with 2 to 7 strains (§4.1.1). This, along with the diversity of the characterized phages, allowed us to provide the first comprehensive analysis of the interactions between *K. pneumoniae* and environmental phages.

Spot tests represent a fast method to initially test phage-bacteria interactions [54,71,112]. We found that <2% of the 6319 unique phage-host combinations tested produced a detectable spot. The sparsity of the interactions is consistent with results observed with other *Klebsiella* lytic phages [59,171,193] or with induced temperate prophages [173]. The definition of phages as having a narrow or broad host-range is very ambiguous [87]. Based on the overall number of interactions and susceptible CLTs, we classified the majority of phages as having a narrow host-range while four were classified as broad-range phages (§4.2.1). Other encapsulated species, such as *E. coli* [306] or *S. enterica* sv. Typhimurium [307], seem to be more prone to phage cross-infection. For instance, in a study with *Salmonella*, phages were termed narrow if they infected dozens of strains, while here they are referred to as broad host-range phages. This suggests that host capsular diversity might determine the degree of modularity of phages in encapsulated bacteria, with *Klebsiella* being at one end of the spectrum. S8/S9 phages escape this capsular modularity and might build a nested phage-bacteria network. Their still narrow host range is consistent with nestedness at the whole network being unlikely [119,308].

It should be noted that a larger panel of hosts for phage isolation might have increased the number of susceptible bacteria. This is especially relevant in CLTs for which we observed no interactions. However, the ability of phages to infect other bacterial strains/CLTs apart from their original host should not vary considerably (given the assumption of 1 depolymerase enzyme: 1 susceptible CLT). The isolation of phages with alternate procedures could reveal a greater diversity of phages shaping the matrix structure [41]. For example, recently Lourenço *et al*. used acapsular *Klebsiella* strains as bait for broad host-range phages [207]. To overcome isolation bias, methods that do not rely on formation of plaques such as viral tagging [124] or cell-free systems [309] might be used.

5.2.1 Predicting the tropism of narrow-range phages

In this work, we first focused on the tropism of narrow host-range phages. Based on the high specificity of phage interactions observed, it was tempting to speculate that they were due to the recognition of capsules by depolymerases. As such, we found that narrow-range phages had a probability of >90% of spotting a bacterium of a specific capsular type if the phage spotted in other bacteria with the same capsule (§4.2.2). Likewise, we found that phages with shared depolymerase domains had a probability of ~60% of spotting a common capsular type (§4.2.3).

To predict phage-bacteria interactions from capsule and depolymerase sequences, the first step is to obtain the **bacterial capsular type**, which can be accomplished by bioinformatic tools [29]. For example, we showed that capsular types inferred with Kaptive [156] correlated very well with the tropism of *Klebsiella* phages. Recently, this tool has been extended to *A. baumannii* [156] and, given the similarities in depolymerase dosage, HGT, and specificity between *Klebsiella* and *Acinetobacter* phages [168], our findings might be also applicable to *A. baumannii*. Other tools, such as SeroBA [310] or hicap [311], are available for *in silico* capsular serotyping of *Streptococcus pneumoniae* and *Haemophilus influenzae*, respectively. However, classification schemes and tools are still needed for most encapsulated bacteria.

Interestingly, there is no tool available to obtain the capsular antigen of *E. coli* or *Salmonella;* instead, current tools determine the O-antigen serotype from

the LPS locus [29,312]. O-antigen-specific depolymerases also exist [164]. Although we have shown here that the O-antigen has no predictive power to detect phage-bacteria interactions in *Klebsiella*, this might not be the case in *E. coli or Salmonella*, where the diversity of the O-antigen is higher and many phages infect in an LPS-dependent fashion [29,312,313]. Therefore, it would be interesting to extend our approach to study interactions between depolymerases and the LPS in species such as *E. coli* or *Salmonella*.

Despite the great potential of *in silico* serotyping, there are several factors that may pose a risk to the successful prediction of a given phage-capsule interaction:

1. Many bacterial strains are untypeable, that is, a capsular type cannot be inferred. For example, in a recent study, 2,419 of 17,752 *Klebsiella* genomes (13.6%) had low or no confident CLT assigned [156]. Similarly, we excluded 10% of the 1,144 genomes analyzed due to low confidence in the serotyping (§4.1.1). Resequencing or long-read sequencing techniques could be used to solve a sizable portion of cases that might have arisen from fragmented genomic assemblies. However, other cases might represent new capsular types that have yet to be identified. This is also supported by rarefaction curves of capsular types which do not show saturation [159].

2. The correlation between capsular genotype and phenotype may not be perfect. For example, single mutations at the capsular locus might modify the specificity of enzymes, resulting in a different serotype. Alternatively, mutations can lead to protein truncation, which can result in capsule inactivation. For example, Haudiquet *et al*. inferred that 3.5% of the *Klebsiella* strains are non-capsulated [149]. We detected at least 4 wild-type strains in our collection (≥ 2.9%) with exclusively translucent colonies, indicative of capsule inactivation [287]. Despite the fact that these bacteria had a capsular type compatible with some phages, viruses were not adsorbed (results not shown). Further efforts should be made to include the prediction of host capsular status (truncated/intact) from genomic analyses.

3. Capsule systems are prone to phase variation, i.e., reversible phenotypic switching in the encapsulated state of bacteria. This process is well-known in *Neisseria meningitidis* [314], *S. pneumoniae* [315], and Bacteroidetes [316]. For example, phase-variable capsular polysaccharides of *B. thetaiotaomicron*, which encode between 2 and 13 different CPS, affect the pattern of phage susceptibility [184]. Interestingly, *Klebsiella* strains of ST307 have an additional CPS system [317]. Whether this system is functional, as well as the contribution of capsular phase-variation in this species, is unknown.

In addition to determining the capsular type of the host, it is also important to determine the **capsular tropism of phages**. Currently, there is no tool available for determining capsular specificity based solely on phage genome sequences. Instead, the typical procedure involves using bioinformatic methods to detect tail spike proteins, followed by the expression and purification of candidate enzymes, and testing their host-range against different capsular types [171]. Given the diversity of phage genomes, this process seems to be applicable only to individual phages. However, we found similar depolymerases between phages with overlapping capsular specificity, revealing a sizable diversity of depolymerases for each capsular type. On the other hand, we found that closely related phages showed little overlapping of capsular specificity. For example, phages from the same similarity group had an 18% probability of infecting the same capsular type (§4.2.4). This observation questions the procedures used to infer phage tropism based on overall genomic similarity with other phages. Instead, the procedure should involve the comparison of receptor-binding domains only, at least within a bacterial species.

Effective switching of tail spike domains between phages has been shown under laboratory conditions [172] or in individual phages [318]. Here, we showed that depolymerase domain swapping is frequent in nature, even across viral families. As a result, phage taxonomy was a poor predictor of host tropism, in contrast to some previous findings in *Ackermanviridae* or *Salmonella* phages [319,320]. Nevertheless, this agrees with the results of phages infecting *Staphylococcus* and *Vibrio* in which overlapping in phage host-range was low

among related phages [54,301]. There are several potential explanations for these results:

i) Some phage families may be more prone to genetic recombination than others [49]. This may be influenced by the presence of recombinases in phage genomes but also by the ecology of the host bacteria. For example, the number of prophages of a given host can affect phage recombination. It was thought that the depolymerase sequences of temperate phages exhibited limited identity with those of lytic phages [173], but our analysis has revealed extensive sharing of depolymerase domains between both types of phages (§4.2.4). This suggests that temperate phages could be an important source of depolymerase domains for lytic phages. Additionally, the presence of specific defense mechanisms of bacteria that interfere with the phage DNA can also affect the recombination rate of phages [54] (**Figure 5.1**).

ii) Overestimation of shared host tropism among related phages could be the result of considering the identity of the complete tail spike. We addressed this by removing the conserved anchor domain, which showed a strong vertical inheritance, and has no role in specificity. In addition, truncation of the depolymerase domain was frequent (§4.2.2), which can lead to comparing the structural domains only. Previous works established a criterion of 200 aa [163,260] to consider a RBP as functional, which led us to include 16 false positive Dpos (data not shown). Consequently, we suggest that a TSP-Dpo must contain at least 500 aa (excluding soluble proteins).

Overall, given the significance of recombination and truncation of TSPs, future studies should address this. For example, determining if depolymerase domain swapping is equally frequent in different phage genera. Additionally, programs for RBPs detection could be more stringent with shorter proteins. The use of programs like AlphaFold [264], which models proteins' tertiary structures from primary amino acid sequences, will also aid in confirming the presence of *bona fide* depolymerases in upcoming works, as shown here (§4.2.2).

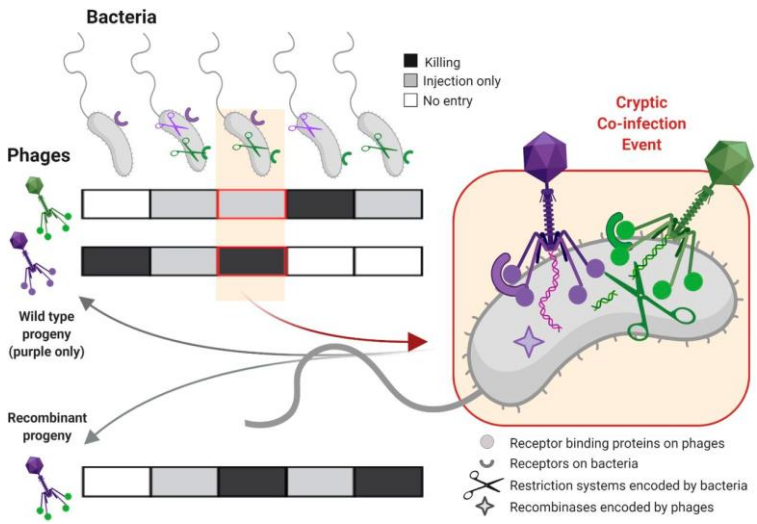

Figure 5.1. Model of cryptic co-infections. Diverse phages co-infect the same host cell using different receptors and receptor-binding proteins. The lytic cycle of a phage might be interrupted by intracellular host defenses. For example, if bacteria have a restriction-modification system that recognizes motifs in the phage DNA. This can result in partially-degraded genomes that can be substrates of recombination. Source: [54].

Depolymerase detection should be coupled with the determination of its specificity. Our results hint at this possibility by using a comparative approach with phages with characterized capsular tropism. Capsular tropism could be determined by host-range assays, but it can also be inferred from past infection events of temperate viruses. For example, to test the ability of depolymerase domains sequences to predict the capsular tropism of other phages we used prophage sequences and the capsular type of the lysogenized host, obtaining significant accuracies for 13 capsular types (KL2, KL3, KL13, KL23, KL25, KL30, KL35, KL47, KL57, KL63, KL64, KL106, and KL140). These correspond to around one-third (n = 2,563 sequences) of total CLT abundance in the global dataset of *K. pneumoniae (https://kleborate.erc.monash.edu/*, n = 8,366 sequences*)*. Surprisingly, we could extend our predictions to *E. coli* strains. These bacteria could have acquired the CPS locus from *Klebsiella*, a process that is common [158]. Therefore, it would be interesting to determine whether phages could also infect *E. coli* strains if there is a compatible capsular type. We tried to find *E. coli* strains to test this hypothesis but failed to find any.

The use of prophage RBD sequences to predict capsular tropism has several limitations: i) the host could undergo capsule swapping after prophage integration [149], leading to false associations between the phage RBD sequence and the observed host CLT (**Figure 5.2.A**); ii) establishing actual RBD-CLT associations can be complex in prophages containing multiple RBDs (**Figure 5.2.B**); and iii) some CLTs can be cross-reactive [166]. This is especially relevant for CLTs in which the sugar composition is not known [192], preventing the comparison of atomic structures as we did in a previous work [59]. Overall, our results hint at the possibility of predicting phage capsular tropism only from depolymerase sequence information, a goal that we achieved partially and that should be further explored in the future.

To overcome the aforementioned limitations, several strategies could be considered. For example, to overcome i), host phylogenies and the presence/absence pattern of certain prophages could be used to identify reliable associations. Briefly, one could estimate the branch on which a prophage was gained [149]. Then, the CLT of all the descendant bacteria could be determined to infer cases of capsule swapping where the prophage-capsular association is not reliable and discard those. Given the high number of prophages in *Klebsiella* [173], restricting the analysis to a subset of all possible prophage-host pairs should not preclude making such associations. For example, we could not assign a consensus CLT for 10 RBD clusters (**Figure 4.15**). Applying this approach, the capsular specificity of these RBDs might be predicted.

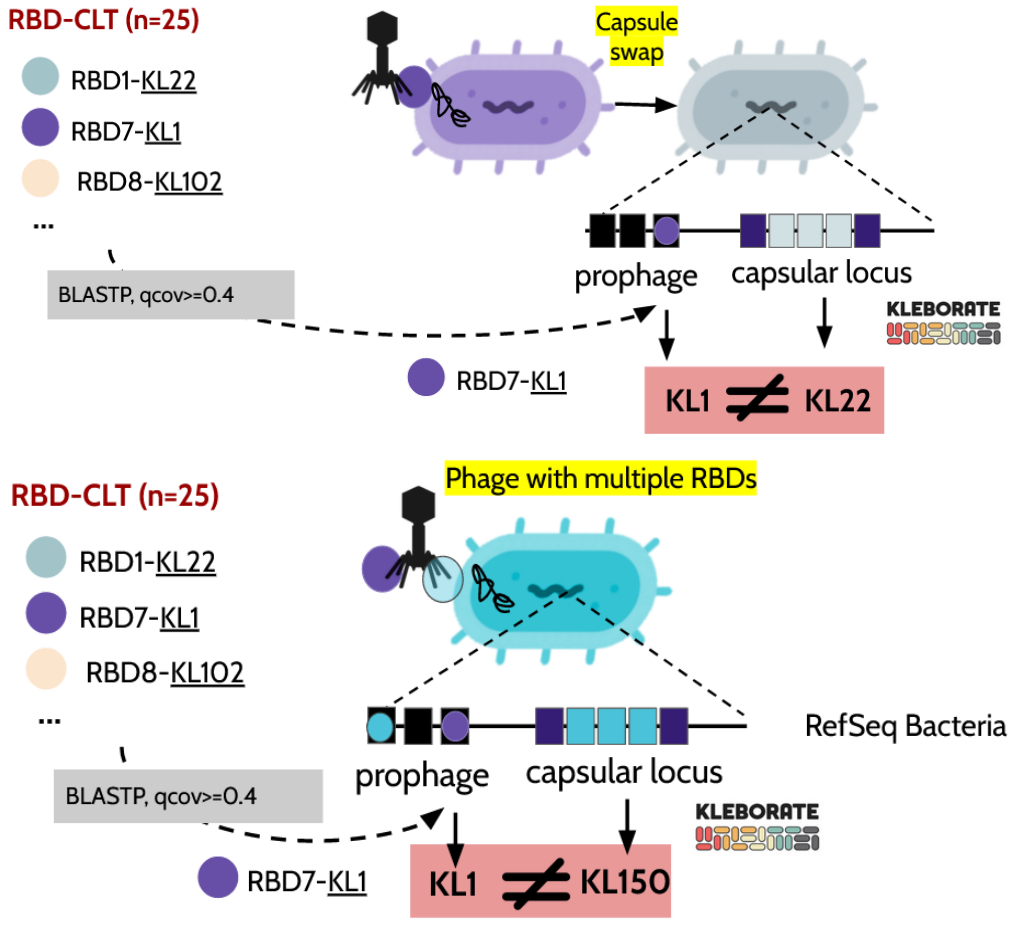

Figure 5.2. Processes that can lead to a false association between a CLT and a RBD when using prophages. A. We would expect that a temperate phage containing an homolog of RBD7-KL1 would be able to lysogenize bacteria from KL1, resulting in a lysogen of KL1 (in purple). However, if the host suffered a capsule swap after phage integration, for example to KL22 (in light green), we would observe a false association between the RBD7 and KL22. **B**. Some phages possess several RBDs and inferring the one that was used to infect a specific capsular type can be challenging. In this example, the temperate phage carried two RBDs, active against KL1 and KL150 (in purple and blue, respectively) but only used the latter to infect a host from KL150. Therefore, a false association would be made between RBD7 and KL150.

5.2.2 From tropism to productive infections

Spot assays do not reflect productive infections. Instead, they can show "lysis from without", in which the bacteria are lysed due to the simultaneous adsorption of virions, or abortive infections [71]. In a previous work, we showed that spots can also result from phage depolymerization activity [59]. To quantify the extent to which capsules and depolymerases are predictive of productive infection by phages, we verified each interaction of the spot assay. We assumed phage replication based on a significant bactericidal effect of phages in bacterial growth curves. Additionally, progeny assays were performed to discard false positives of this killing assay. We did not find productive infection in 30 of 124 positive phage-bacteria interactions (24%). This false positive rate of spot assays as predictors of productive infections is much higher than the 6% observed in a previous study using *Salmonella* as host [271]. However, in another study with *Klebsiella*, the rate of false spots was around 50% [321]. This difference might be due to the prevalence of depolymerases in *Klebsiella* phages, in which the enzymatic activity would result in a lighter bacterial lawn at the site of phage application and, therefore, a positive spot.

To test whether successful depolymerization occurred in cases of phage avirulence, we developed a liquid depolymerization assay as an alternative to capsule staining and microscopy, which is a laborious procedure and does not produce quantitative data [59]. This assay is based on indirectly detecting the depolymerase activity of phages when the cells are in a stationary phase and most phages cannot replicate [322]. Depolymerase activity was quantified as the reduction in capsular carbohydrates of phage-treated bacteria compared to untreated controls. A few considerations about the depolymerization assay are the following:

- It was essential to wash the cells before proceeding to capsule extraction. Otherwise, differences with untreated bacteria were not observed. This is probably due to the interference of capsule debris in the supernatant after phage depolymerization activity. As it will dramatically reduce the time of this assay, it would be interesting to test whether the quantification of the

supernatants of untreated and treated bacteria is sufficient to reveal depolymerization activity.

- The initial yield of encapsulated cells is of great importance. Some bacterial strains expressed low levels of capsular polysaccharides, close to the assay's detection limit. To overcome this, we increased the overnight culture volumes from 0.5 mL to 2 mL and added an incubation step of two hours at 56 ºC to resuspend carbohydrates in water before quantification [155]. However, we were still unable to test two phage-bacteria combinations as the wild-type bacteria rapidly became non-encapsulated *in vitro*. To improve this, we could try shortening the co-incubation period from overnight to a few hours or use a minimal medium, where bacterial capsule inactivation is rare [155]. Alternatively, exponential cells could be used because capsule expression is higher at this phase [323]. Incubation in saline buffer, at lower temperatures, or in the presence of bacteriostatic antibiotics could be done to avoid bacterial replication in the presence of phage and, therefore, capsule restoration.

Despite the limitations of the depolymerization assay, we were able to determine that narrow-range phages produced capsule digestion in every host in which they produced a spot without virulent infection. Therefore, we confirmed the reliability of spot assays to determine capsule-dependent tropism and phage-induced depolymerization in the absence of virulent infections.

To better understand avirulent infections, we performed adsorption assays. All the phages except one, exhibited high adsorption rates, suggesting the presence of functional receptors. Although it is well-known that the phage productive host range is narrower than the adsorptive host range, this is rarely quantified. For instance, *Vibrio* phages effectively attach to nearly-clonal bacteria; however, mobile defense mechanisms block infection [112,113]. Our results show that post-adsorption resistance hindered the predictability of phage infection based on capsules, leading to a drop in sensitivity from 92% to 53%. We classified productive interactions based on the presence of lytic activity of phages in bacterial growth curves. Further quantitative analysis, such as plaque assays, may reveal

variations in adsorption efficiency due to receptors or even a more prominent role of intracellular defenses.

The role of known or novel intracellular defense mechanisms in the observed phenotypes remains to be characterized. In this work, we selected bacterial strains to maximize the number of capsular types. The next step would be to include more strains with a common capsular type, selecting those with different defense systems and accessory genes. For example, the presence of plasmids could affect phage infectivity [200]. We discarded CRISPR-Cas and superinfection exclusion mechanisms and found no correlation between the total number of defense systems and post-adsorptive resistance [324]. For example, phages S13a and S13c lysed different strains from KL102 with 10-11 defense but not CU618, which only has 5. CU618 is the only strain in the collection with the RexAB system, a prophage-encoded defense mechanism [325]. RexAB, or additional candidate systems (**§4.2.5**), could be knocked down in resistant bacteria or expressed in susceptible hosts. New defense mechanisms can also be discovered through transposon-insertion sequencing (Tn-Seq) or CRISPR interference (CRISPRi) assays [326,327].

Predicting the transductive host-range of phages would be the next major goal. To achieve this, we first need to confirm that the phage DNA is being delivered to the host. One possible way to do this is by measuring changes in potassium levels after infection [328]. Additionally, restriction-modification (RM) systems are present in more than 80% of prokaryotic genomes, making them by far the most prevalent type of defense [98]. As a result, they are considered the main barrier to bacterial genetic modification [329] and, potentially, to phage replication. To overcome this, PacBio sequencing can be used to identify RM motifs in resistant hosts and check for their presence in phage genomes. As an alternative, a method such as the one developed here to extract RBDs and assign CLT specificity could be expanded to RM enzymes based on the specificity domains [330].

Finally, phages´ lifestyles can also be a barrier to productive infections. Our analyses found that phages did not encode integrases (**§4.1.4**). However, some phages can replicate in the cytoplasm, similarly to plasmids [331]. For example, group 13 phages have ParB proteins involved in plasmid partitioning and showed

identity to phage-like plasmid genomes (results not shown). Therefore, phages might be maintained as plasmids in some hosts while in others they could act as strictly lytic due to epistatic interactions. Our work found a high discrepancy between *in silico* lifestyle predictions for phages, even when the confidence score was high (§4.1.4). Therefore, there is an urgent need to improve our understanding and classification of phage lifestyle strategies. With all these considerations, we might be able to predict the different layers of phage-bacteria interactions in the near future [136].

5.2.3 The role of capsules in productive infections

An increasing number of works report capsule inactivation as the main cause of resistance to capsule-specialist phages [91,176,178,183,207], which suggests that capsules are indeed required for infection. To test that in our phage collection, we aimed to isolate acapsular (Cap-) mutants for phage-susceptible bacteria. Capsule inactivation can occur i*n vitro* without phage pressure [155,287]. Here, we have shown that 90% of the *Klebsiella* strains showed heterogeneous Cap+/Cap- phenotypes. One single colony was used to grow bacteria in liquid for two passages before determining Cap+/Cap- mutants. Surprisingly, Cap- mutants overall represented ~ 10% of the bacterial culture. This can be the result of both, a high incidence of capsule-disruption mutations [149,287], and also from selection, as capsule production is costly in rich media [155].

Capsule inactivation provided resistance to narrow-phages in 85% of the combinations tested (§4.3.2). Therefore, the high incidence of capsule-inactivating mutations can affect phage resistance. Additionally, we observed that wild-type bacteria evolved resistance to most phages very fast (~ 6 hours post-infection). This could be explained by pre-existing variation in bacterial cultures, that is, non-encapsulated mutants being quickly selected by phages [332,333].

Unexpectedly, capsule-specialist phages were still able to infect Cap- mutants in 15% of the combinations. Carbohydrates of the capsule are thought to serve as the primary receptor for phages with depolymerases, followed by irreversible attachment to a secondary protein receptor [334] or components of the LPS [82]. Whether the reversible binding to receptors is necessary for the

subsequent infection process has not been studied systematically. A well-studied case is that of phage T4, in which the initial binding to OmpC or the O-antigen is required for further infection [335]. Alternatively, *Drexleviridae* phages are thought to reversibly bind to the O-antigen glycans without requiring this for host recognition [82]. All our 3 *Drexleviridae* phages showed a narrow host-range restricted by the capsular locus type, in agreement with specific external glycans acting as primary receptors. However, while *Klebsiella* phages D7a-b were able to infect some Cap- mutants, phage D7c infected none. This was in line with RBPs found in phages D7a-b, which might recognize proteinaceous receptors (§4.3.2). Apart from *Drexleviridae* phages D7a-b, we found other 8 narrow-phages able to infect Cap- mutants. These phages were distributed across groups (A1, M12, and S13) and singletons phages (S6a and S10a) (§4.3.2). Differences between taxonomically-related phages might reflect distinct RBPs or domains acquired by HGT, similar to the case of depolymerases. This will also prevent making conclusions about the primary and secondary receptors of phages based on their taxonomy.

Besides RBPs, one plausible explanation for infectivity patterns of Cap- mutants is the specific mutations that lead to capsule inactivation. This is further supported by the fact that, within a single phage-bacteria combination, some Cap- mutants were infected while others were not, suggesting different effects of capsule-inactivating mutations. Previous studies have shown that secondary mutations were rare when sequencing spontaneous acapsular variants [287]. However, some mutants might have an altered LPS structure, given the possible pleiotropic effects of mutations affecting surface polysaccharides [336–339]. For example, mutants with altered glucosyltransferases showed an excess production of LPS when the capsule was down-regulated [340]. In this situation, facultative-capsular phages can be misidentified because overproduction of LPS might avoid recognition of secondary receptors [334]. In another study, it was found that LPS O-antigen mutants are capsule deficient due to a reduction in capsule polysaccharide retention [341]. Therefore, we cannot exclude that some of the Cap- mutants included have also defects in the LPS, which could also play an important role as secondary receptors. Additional surface polysaccharides such as the novel glycan exploited by N4-like phages [337] might be affected too.

Independently of the mechanisms that lead to capsule inactivation, our results highlight the importance of studying this process due to its high frequency and impact on phage infection. Capsule inactivation has also been observed *in vivo* in many species [342]. For example, in *K. pneumoniae*, encapsulated and non-encapsulated variants have been isolated from urine, blood, and wound samples of the worrisome ST258 [154]. Overall, it is estimated that more than 3.5% of *K. pneumoniae* sequenced strains are non-capsulated [149] and mutations disrupting capsule biosynthesis genes have been observed in several lineages recurrently, an strong sign of adaptive evolution [149,154]. This parallel evolution of capsular biosynthetic mutations has also been observed in the evolution of *Bacteroides fragilis* in the gut of healthy people [343] or in *B. thetaiotaomicron* in mice with different dietary conditions [344].

Despite reports of capsule inactivation *in vivo*, there is a lack of knowledge about the evolution of non-encapsulated bacteria in certain environments. For example, capsule inactivation by gene disruption could be quantified in metagenomic shotgun samples for bacteria encoding a capsule system [146,149]. Not only stable capsule inactivation, but phase-variable shifts or down-regulation of the capsule, might be underestimated [183,184,337]. Here, we observed a subset of bacterial strains with Cap- revertants, whose mechanisms remain to be further studied.

5.2.4 The pitfalls of host-range prediction for broad-range phages

In previous sections we have discussed the host-range of narrow-range phages, which showed a strong dependency on the capsule due to their RBPs with depolymerase domains. This dependency was first shown when the capsular type of the host predicted >90% of narrow-range phage interactions. On the other hand, the capsular type predicted ~40% of the interactions for broad host-range phages belonging to groups S8/S9 (**Figure 4.12**). This sensitivity was far from random, suggesting that broad-range phages have a certain degree of capsular tropism. However, these phages did not show strong depolymerization activity or encoded canonical depolymerase domains (**§4.2.6**). Instead, S8 phages have several

carbohydrate-binding modules, which might allow them to bind capsules and approach the cell surface. Prophage analyses similar to those performed with narrow-range phages showed that these domains are widespread across capsular types (results not shown), suggesting that they might be promiscuous. Then, they might use their long, flexible tails to find secondary receptors, as seen in phages infecting Gram-positive bacteria [345,346]. Hence, these siphophages might be able to overcome the capsule without the need for a catalytic-degrading activity. To check this, carbohydrate-binding modules of RBPs can be fused to a fluorescent marker and cell binding can be determined by fluorescence-activated cell sorting (FACS) [124] or microscopy [347]. Additionally, we found that the S9a phage had an esterase domain (§4.3.2) which can deacetylate surface polysaccharides but leave the backbone of the polysaccharide intact [174].

The partial capsular dependency of broad host-range phages was also shown with the Cap- mutants infectivity assays (§4.3.2). S8 phages showed a polyvalent mode of action, that is, they required the capsule for infecting a subset of bacterial strains, while for others they were able to infect acapsular bacteria. Therefore, the broad-range of S8 phages was not the result of recognizing a conserved bacterial receptor but, instead, of using different strategies. On the other hand, phage S9a could not infect any Cap- mutant, confirming the requirement of capsule deacetylation to initiate infection [174]. Importantly, acetylation variation could be observed within a capsular type. For example, some KL57 strains present a truncated gene encoding for the acetyltransferase [348]. Capsule acetylation determinants could also be encoded in prophages [349,350], adding an additional layer of complexity.

Given their ability to infect acapsular bacteria, it can also be speculated that S8 phages might preferentially encounter receptors in the equator of the cell surface, where the capsule is thinner [351], or in bacteria with lower capsule production levels. For example, some *Shigella* phages showed a preference for low EPS-producing bacterial hosts [169]. We found that S8 spots that did not yield productive infections were very turbid, probably due to the infection of a subset of non-encapsulated variants in the population (Figure 4.20). This was further supported by the fact that S8 phages preferentially infected bacteria with a higher prevalence of non-encapsulated variants (Figure 4.22). Additionally, S8 phages

showed the largest stochasticity in spotting assays (§4.2.1). All these results suggest that heterogeneity in bacterial populations affects the infection phenotype of these phages. Supporting this, *Klebsiella* phage mtp5, which exclusively infects non-encapsulated variants *in vitro*, was able to replicate in the mouse gut colonized with wild-type bacteria [207]. This further suggests that non-encapsulated variants emerge in natural environments that can then be infected by broad host-range phages. If acapsular variants represent a small and transient population, this supports the selection of more generalist phages [352], which are likely underrepresented in this study.

Further work will be required to completely understand the tropism of S8 polyvalent phages. Isolation of phage-resistant clones and mutation identification is the typical procedure to obtain candidate receptors [176,353]. Contrary to phage-resistant clones of encapsulated bacteria, phage-resistant clones of acapsular mutants showed mutations in a variety of structures such as the O-antigen, LPS, and genes encoding protein transporters [207]. The use of proteinaceous receptors might explain the dynamics that we observed in liquid, where resistance is delayed over capsule-obligate interactions (§4.3.4). Mutations in single genes will be less frequent than in complete operons and proteins are typically more functionally constrained than carbohydrates [29]. For example, *Escherichia* phage U136B uses TolC, an outer membrane protein, and the core of the LPS for infection, with LPS-related mutations occurring in 70% of cases [354]. Additionally, the fitness cost could also be higher. For instance, reduced porin expression has a high fitness cost due to the exclusion of nutrients from the periplasm [355].

5.3 Phages in natural communities

Bacteria and phages are not isolated entities but are found in complex microbial communities. Hence, the interactions of local pairs of phage-bacteria will affect the evolutionary response of other members of the community [356]. In this regard, establishing the role of bacterial capsules for different phages is not trivial, as it would affect the co-evolutionary dynamics of both (bacteria and phages) but also of the other members in the community. Most phages of this work were isolated from a common sample, or close location (**Figure 4.4**) and, in consequence, they are likely to co-exist in the natural environment.

First, we showed that the capsular tropism of narrow-range phages was mainly explained by the horizontal acquisition of Dpos domains in TSPs. This would be an example of arms race dynamics (ARD), in which phages increase their infectivity without losing the ability to infect the ancestral capsular type (**§1.2.4**). The fact that many phages appear with truncated TSPs supports this, providing an anchoring point to the acquisition of novel Dpos (**Figure 4.13**) and, therefore, novel specificities. Nevertheless, the physical limits of phages to carry additional TSPs will pose a constraint on the specificities that a phage can evolve. On the other side, we observed that bacteria´s intracellular defenses might block the infection of phages (**§4.2.6**). The rapid acquisition of bacteria defense systems, in the so-called defense-islands [357], would be another example of ARD. Acquiring a novel defense mechanism is not expected to impact resistance to ancestral phages. Instead, synergism between defense systems is common [358]. Still, fitness costs will probably impact the evolution of general resistance of bacteria [118].

Alternatively, phages do not necessarily go through an intermediate step of regressive evolution, such as the truncation of a RBP but, instead, might directly swap its specificity module [163]. This would be an example of fluctuating selection dynamics (FSD), in which the phage would gain a new specificity but lose its tropism for the ancestral capsular type. Concomitantly, *Klebsiella* can undergo capsule swaps, gaining resistance to a subset of phages, but at the cost of being susceptible to other capsular-specific phages. In this scenario, frequency-dependent selection may act, leading to cycles of phage depolymerase domains and bacterial capsule swapping.

The fact that phages from the same location appear with different TSP specificities does not support that environments are dominated by a single capsular type of *Klebsiella*. Instead, our results suggest that a subset of phages have specialized in infecting different capsular types. Likewise, both broad- and narrow-range phages were isolated from the same sample. Recently, Lourenço *et al*. isolated phages that exclusively infected encapsulated or non-encapsulated *K. pneumoniae* strains from the river Seine and sewage [207]. These results, taken together, suggest a coexistence of not only different serotypes but also encapsulated and non-capsulated bacteria in natural environments. Still, longitudinal studies of *K. pneumoniae* and its phages in the natural environment are lacking.

Identification of the receptor(s) for each phage will allow understanding evolutionary pressures. Here, we have addressed whether the CPS was an obligate receptor by challenging acapsular bacteria with phages. Given the importance of the CPS in bacterial ecology, evolution, and pathogenesis [29], phage resistance mediated by CPS inactivation could have important fitness consequences [91,182]. Additionally, we found that bacterial resistance to phages is negatively correlated with their capsule essentiality. As such, resistance to capsule-obligate phages is higher than to facultative-capsular phages, in line with recent results [183,207]. This could be due to the large size of the capsular operon subjected to mutation, HGT, and recombination. As such, this would be an example of asymmetry in phage and bacteria co-evolution [359]. While there are numerous mutational pathways for bacteria to modify or lose their CPS, phages would need to acquire new RBP specificities, which are more constrained. This might cause capsule-obligate phages to reach an evolutionary dead-end and lose their capacity to co-evolve.

Finally, laboratory experiments might not reflect infection dynamics in natural environments in which diversity, fitness costs, and population structure will considerably vary. First, while only a subset of phages and bacteria are tested in the laboratory, co-evolution in nature will be affected by the genetic composition of the whole population. For example, inactivation of the capsule may come at the cost to the cell of cross-sensitization to facultative-capsular phages [91,178]. Alternatively, phages might overcome bacterial resistance by the acquisition of new RBPs by HGT with other phages. The existence of (poly)lysogenic bacteria

and temperate phages will be another layer of complexity [183]. On the other hand, inactivation of the CPS might be more constrained *in vivo* given its importance in different host- and non-host -associated environments [147,154,343]. Here, we used a rich, well-mixed medium to study phage-dynamics, that, in turn, favors capsule inactivation [287]. Contrarily, heterogeneous environments can create permanent or transient spatial refuges for bacteria by limiting phage dispersal and allowing stable coexistence [360,361]. While controlled laboratory experiments pave the road to better understand the arms race between bacteria and phages, more complex, community-aware experiments will be necessary to increase our understanding of phage and bacteria co-evolution in nature.

5.4 Applications

In recent years, phages have re-emerged as a research topic of significant interest. This has been motivated by the increased understanding of phage diversity, their role in shaping microbial communities, and their importance in health and disease. This renaissance of phage research has also been driven by the urgent need for alternative therapies to treat multidrug-resistant bacteria. Shortly after the discovery of phages by Frederick Twort in 1915 [1], Felix d'Herelle proposed their use to treat bacterial infections [2], which is referred to as phage therapy. However, with the discovery, effectiveness, and easy production of antibiotics, phages were almost forgotten. The exceptions were the Eliava Institute of Bacteriophage, Microbiology, and Virology, in Georgia and the Ludwik Hirszfeld Institute of Immunology and Experimental Therapy, in Poland, where most of our knowledge about phage therapy comes from. In Western Europe and the USA, studies under the guidelines of the FDA or EMA have been restricted until recently [362]. Therefore, although some phage-based products are available in some Eastern European countries, there are no phage products approved for human therapy in the USA or the rest of Europe to date [363], with the exception of cases of compassionate treatment [364,365].

Today, antibiotic-resistant infections cause at least 1.3 million deaths per year. These numbers are estimated to increase by up to 10 million deaths by 2050 if no urgent actions are taken. Concretely, the World Health Organization (WHO)

has claimed the urgent need for alternative therapies against ESKAPE bacteria (*Enterococcus faecium*, *Staphylococcus aureus*, *Klebsiella pneumoniae*, *Acinetobacter baumannii*, *Pseudomonas aeruginosa*, and *Enterobacter* spp.), which causes more than 70% of deaths due to antimicrobial resistance [366]. In this work, we have focused on one ESKAPE pathogen, *K. pneumoniae*, simulating a scenario in which environmental phages might be used to target clinical strains. In this regard, the results of this thesis are relevant for phage therapy of *K. pneumoniae*.

First, we found that the probability of a *Klebsiella* phage cross-infecting another bacterium from its original host is quite low (**Figure 4.11**). This might compromise phage therapy as a highly personalized medicine in *K. pneumoniae*. This is not the case in other species such as *S. aureus* in which a few phages can target many bacterial strains [301,362]. Importantly, we showed that most phage-bacteria interactions in *K. pneumoniae* are predictable by the capsular type and depolymerases. For eight CLTs (KL20, KL126, KL134, KL139, KL140, KL149, KL151 and KL157) no phages have been characterized previously, increasing the availability of phages that target diverse capsular types. This presents a "*sur mesure*" approach to phage therapy, in which phage collections are established, characterized, and prepared for use in clinical situations [367].

Second, we found that the Dpos and CLT associations were not predictable of productive infections due to bacterial intracellular defenses. Given the arsenal of bacterial defense mechanisms and their diversity across closely related strains, this poses a risk to the success of phage therapy. Even the use of phages able to complete their full-life cycle for therapy is desirable, establishing solid associations between RBDs and pre-adsorptive phage activity might be useful. If the bacterial CLT is known, this would aid the selection of phages without the need for empirical testing. In addition, even if the phage is unable to complete its cycle, its depolymerization activity may aid in the removal of infection by the immune system or antibiotics [91,182]. Finally, this can help to engineer previously characterized phages, able to circumvent intracellular defenses, with new specificities. The modular structure of depolymerases makes this possible, and it has been demonstrated to work in *Klebsiella* phages [172]. Given the legislative hurdles to the use of engineered phages [368], an alternative approach that holds promise is the direct use of depolymerases [369–371].

Third, the development of phage resistance is likely an unavoidable consequence of phage therapy. Here, we have addressed if the recurrent phenomenon of capsule loss leads to phage resistance. In order to slow down the emergence of resistance, the use of cocktails of phages with alternate mechanisms of action has been proposed [178,372]. Whilst determining the host range of a phage can be easily done in the laboratory, identifying the specific cell surface receptor is more challenging [71,86]. However, the identification of the approximate receptor function/location might be enough in some scenarios. For instance, to design a phage cocktail it might not be necessary to determine the specific carbohydrate recognized by the phage as simply knowing that the phage attaches to the capsule would suffice [86]. This could also help to rationalize the combinatorial use of phages and antibiotics, known as phage–antibiotic synergy (PAS). For example, capsule loss has been associated with increased susceptibility to beta-lactam antibiotics in *A. baumannii* [91]. In contrast, bacterial hypermucoidity, often associated with capsular hyperproduction [373], could result in increased susceptibility to antibiotics [147].

Overall, a better understanding of phage-bacteria interactions will help to rationalize phage therapeutics and, therefore, contribute to their success.

5.5 Personal reflection and final remarks

Phage research is moving very fast and during the development of this thesis major breakthroughs have been made. Large collections of bacteria and phages have been made available [54,82,113,301,374]. The diversity of phages continues to surprise us: from the use of alternative codons [375], and modified nucleotides [376], to megaphages [35], and the RNA virome [377]. Concomitantly, the number of bacterial strategies to avoid being killed by these parasites is also outstanding: from L-form conversion in Gram-positive bacteria [378], to chemical defense [379], or defense systems with parallelism with eukaryotic immunity [380,381]. Not only bacterial parasites but parasites of phages, so-called phage satellites, are in the spotlight [382,383].

All of these discoveries have been made possible thanks to addressing technological gaps in high-throughput sequencing, metagenomics, and

bioinformatics, among others. For example, during the development of this work, no tools were available for consistently annotating phage genomes [250], detecting phage receptor-binding proteins [260], or phage defense systems [98,384]. Major breakthroughs are still on hold. Phage comparative genomics has been limited by the lack of a conserved marker among phages [49]. The recent development of algorithms able to predict protein structures from primary amino acid sequences [264] opens the door to new questions that a few years ago could not be even framed. Additionally, advances in culturomics and culture-independent approaches will shed light on the dark microbial matter.

The COVID-19 pandemic has had a profound impact on the world, bringing significant changes to many aspects of society, including science. The sudden need for new technologies and to combat a viral disease with no available treatment has accelerated research and innovation in multiple areas. Antibiotic resistance is considered a silent pandemic, as the number of bacterial infections no longer treatable is increasing sharply worldwide. Phage therapy and phage-derived products hold the promise to treat bacterial infections, and improve bacteria-associated syndromes such as inflammatory bowel diseases [385], or diagnosis [386]. To make these promises a reality, applied research needs to be coupled with understanding the fundamental mechanisms underlying the interactions between phages and bacteria. I hope this work has contributed towards this goal.

As a final remark, the COVID-19 pandemic has also reflected the importance of globalization and highlighted that we live in a highly interconnected world, full of inequities. Science has a profound responsibility within society, and as such, it should not only provide accurate, ethical, and open research, but also ensure that everyone could benefit from the results obtained. During the development of this work, I was very lucky to supervise students, teach at the university, and engage in multiple outreach activities about the importance of the microbial world and evolution. While the results of these activities are not explicitly reflected in this thesis, I realized that the joy of a talk at a conference or a paper is nothing compared to the enthusiastic face of kids engaged with your research. In times of “publish or perish” culture, my hope is that the future years will not only bring groundbreaking discoveries, but also a generation of scientists compromised with their peers, society, and equality.

6. Conclusions

Phage-bacteria interactions shape a myriad of ecological processes which are relevant for explaining the diversity and evolution of microorganisms, but also for understanding health and disease. In this work, we aimed at gaining a deeper insight into the determinants of phage-bacteria interactions, with an emphasis on the interplay between capsules and depolymerases. The main conclusions of this thesis are the following:

1. The Kpn-phage is a representative collection of *K. pneumoniae* and associated phages to study the determinants of phage-host interactions.
2. The capsular diversity of *Klebsiella* restricts the host range of most phages, being not only the main factor of specificity but also a requirement for infection.
3. Phage-encoded depolymerase domains can predict capsular tropism, which can be inferred using comparative approaches and prophage datasets.
4. The predictability of capsular tropism is limited by post-adsorptive resistance mechanisms different from prophages or CRISPR-Cas.
5. Phages lacking depolymerases exhibit broader host ranges. This is the result of a polyvalent mode of action as S8 phages exhibited both capsule-dependent, likely mediated by carbohydrate-binding-modules, and capsule-independent tropism.
6. The frequency and contingency of capsule-inactivating mutations affect the phage host range and both narrow- and broad-range phages can infect acapsular bacteria.
7. The role of the capsule as an obligate or facultative receptor affects phage-bacteria dynamics in terms of resistance development and its cost but not in virulence.
8. Overall, comprehensive datasets combining phenotypic and genomic information, similar to the data presented here, are necessary to improve our understanding and the predictability of phage-bacteria interactions in nature and also for applications of phages.

7. References

1. Twort FW. An investigation on the nature of ultra-microscopic viruses. The Lancet. 1915. pp. 1241–1243. doi:10.1016/s0140-6736(01)20383-3

2. Herelle F d'. The bacteriophage: its role in immunity. 1921.

3. Suttle CA. Viruses in the sea. Nature. 2005. pp. 356–361. doi:10.1038/nature04160

4. Comeau AM, Hatfull GF, Krisch HM, Lindell D, Mann NH, Prangishvili D. Exploring the prokaryotic virosphere. Research in Microbiology. 2008. pp. 306–313. doi:10.1016/j.resmic.2008.05.001

5. Dion MB, Oechslin F, Moineau S. Phage diversity, genomics and phylogeny. Nat Rev Microbiol. 2020;18: 125–138.

6. Weinbauer MG. Ecology of prokaryotic viruses. FEMS Microbiology Reviews. 2004. pp. 127–181. doi:10.1016/j.femsre.2003.08.001

7. Weinbauer MG, Rassoulzadegan F. Are viruses driving microbial diversification and diversity? Environ Microbiol. 2004;6: 1–11.

8. Kim M-S, Bae J-W. Lysogeny is prevalent and widely distributed in the murine gut microbiota. The ISME Journal. 2018. pp. 1127–1141. doi:10.1038/s41396-018-0061-9

9. Barr JJ, Auro R, Furlan M, Whiteson KL, Erb ML, Pogliano J, *et al*. Bacteriophage adhering to mucus provide a non–host-derived immunity. Proceedings of the National Academy of Sciences. 2013. pp. 10771–10776. doi:10.1073/pnas.1305923110

10. Fortier L-C, Sekulovic O. Importance of prophages to evolution and virulence of bacterial pathogens. Virulence. 2013;4: 354–365.

11. Połaska M, Institute of Agricultural and Food Biotechnology, Department of Microbiology, Rakowiecka 36, Warsaw 02-532, Poland, *et al*. Bacteriophages—a new hope or a huge problem in the food industry. AIMS Microbiology. 2019. pp. 324–346. doi:10.3934/microbiol.2019.4.324

12. Alsaadi A, Beamud B, Easwaran M, Abdelrahman F, El-Shibiny A, Alghoribi MF, *et al*. Learning From Mistakes: The Role of Phages in Pandemics. Front Microbiol. 2021;12: 653107.

13. Gordillo Altamirano FL, Barr JJ. Phage Therapy in the Postantibiotic Era. Clin Microbiol Rev. 2019;32. doi:10.1128/CMR.00066-18

14. Wigington CH, Sonderegger D, Brussaard CPD, Buchan A, Finke JF,

Fuhrman JA, *et al*. Re-examination of the relationship between marine virus and microbial cell abundances. Nat Microbiol. 2016;1: 15024.

15. Williamson KE, Fuhrmann JJ, Wommack KE, Radosevich M. Viruses in Soil Ecosystems: An Unknown Quantity Within an Unexplored Territory. Annu Rev Virol. 2017;4: 201–219.

16. Hoyles L, McCartney AL, Neve H, Gibson GR, Sanderson JD, Heller KJ, *et al*. Characterization of virus-like particles associated with the human faecal and caecal microbiota. Res Microbiol. 2014;165: 803–812.

17. Howard-Varona C, Hargreaves KR, Abedon ST, Sullivan MB. Lysogeny in nature: mechanisms, impact and ecology of temperate phages. The ISME Journal. 2017. pp. 1511–1520. doi:10.1038/ismej.2017.16

18. Ackermann H-W. 5500 Phages examined in the electron microscope. Archives of Virology. 2007. pp. 227–243. doi:10.1007/s00705-006-0849-1

19. Hyman P, Abedon ST. Smaller Fleas: Viruses of Microorganisms. Scientifica. 2012. pp. 1–23. doi:10.6064/2012/734023

20. Bradley DE. Ultrastructure of bacteriophage and bacteriocins. Bacteriological Reviews. 1967. pp. 230–314. doi:10.1128/br.31.4.230-314.1967

21. Turner D, Kropinski AM, Adriaenssens EM. A Roadmap for Genome-Based Phage Taxonomy. Viruses. 2021;13. doi:10.3390/v13030506

22. Nobrega FL, Vlot M, de Jonge PA, Dreesens LL, Beaumont HJE, Lavigne R, *et al*. Targeting mechanisms of tailed bacteriophages. Nat Rev Microbiol. 2018;16: 760–773.

23. Dennehy JJ, Abedon ST. Adsorption: Phage Acquisition of Bacteria. In: Harper DR, Abedon ST, Burrowes BH, McConville ML, editors. Bacteriophages: Biology, Technology, Therapy. Cham: Springer International Publishing; 2021. pp. 93–117.

24. Hu B, Margolin W, Molineux IJ, Liu J. Structural remodeling of bacteriophage T4 and host membranes during infection initiation. Proceedings of the National Academy of Sciences. 2015. doi:10.1073/pnas.1501064112

25. Mäntynen S, Laanto E, Oksanen HM, Poranen MM, Díaz-Muñoz SL. Black box of phage–bacterium interactions: exploring alternative phage infection strategies. Open Biology. 2021. p. 210188. doi:10.1098/rsob.210188

26. Feiner R, Argov T, Rabinovich L, Sigal N, Borovok I, Herskovits AA. A

new perspective on lysogeny: prophages as active regulatory switches of bacteria. Nat Rev Microbiol. 2015;13: 641–650.

27. Rakonjac J, Bennett NJ, Spagnuolo J, Gagic D, Russel M. Filamentous bacteriophage: biology, phage display and nanotechnology applications. Curr Issues Mol Biol. 2011;13: 51–76.

28. Venturini C, Petrovic Fabijan A, Fajardo Lubian A, Barbirz S, Iredell J. Biological foundations of successful bacteriophage therapy. EMBO Mol Med. 2022;14: e12435.

29. Mostowy RJ, Holt KE. Diversity-Generating Machines: Genetics of Bacterial Sugar-Coating. Trends Microbiol. 2018;26: 1008–1021.

30. Bernheim A, Sorek R. The pan-immune system of bacteria: antiviral defence as a community resource. Nature Reviews Microbiology. 2020. pp. 113–119. doi:10.1038/s41579-019-0278-2

31. Ofir G, Sorek R. Contemporary Phage Biology: From Classic Models to New Insights. Cell. 2018;172: 1260–1270.

32. Penadés JR, Chen J, Quiles-Puchalt N, Carpena N, Novick RP. Bacteriophage-mediated spread of bacterial virulence genes. Curr Opin Microbiol. 2015;23: 171–178.

33. Cenens W, Makumi A, Govers SK, Lavigne R, Aertsen A. Viral Transmission Dynamics at Single-Cell Resolution Reveal Transiently Immune Subpopulations Caused by a Carrier State Association. PLoS Genet. 2015;11: e1005770.

34. Zrelovs N, Dislers A, Kazaks A. Motley Crew: Overview of the Currently Available Phage Diversity. Front Microbiol. 2020;11: 579452.

35. Al-Shayeb B, Sachdeva R, Chen L-X, Ward F, Munk P, Devoto A, *et al*. Clades of huge phages from across Earth's ecosystems. Nature. 2020;578: 425–431.

36. Chaudhari HV, Inamdar MM, Kondabagil K. Scaling relation between genome length and particle size of viruses provides insights into viral life history. iScience. 2021;24: 102452.

37. Hatfull GF, Hendrix RW. Bacteriophages and their genomes. Curr Opin Virol. 2011;1: 298–303.

38. Brüssow H, Hendrix RW. Phage genomics. Cell. 2002;108: 13–16.

39. Cook R, Brown N, Redgwell T, Rihtman B, Barnes M, Clokie M, *et al.* INfrastructure for a PHAge REference Database: Identification of Large-Scale Biases in the Current Collection of Cultured Phage Genomes. Phage (New Rochelle). 2021;2: 214–223.

40. Roux S, Krupovic M, Daly RA, Borges AL, Nayfach S, Schulz F, *et al.* Cryptic inoviruses revealed as pervasive in bacteria and archaea across Earth's biomes. Nature Microbiology. 2019. pp. 1895–1906. doi:10.1038/s41564-019-0510-x

41. Kauffman KM, Hussain FA, Yang J, Arevalo P, Brown JM, Chang WK, *et al.* A major lineage of non-tailed dsDNA viruses as unrecognized killers of marine bacteria. Nature. 2018;554: 118–122.

42. Krishnamurthy SR, Wang D. Origins and challenges of viral dark matter. Virus Res. 2017;239: 136–142.

43. Paul JH, Sullivan MB. Marine phage genomics: what have we learned? Curr Opin Biotechnol. 2005;16: 299–307.

44. Breitbart M, Thompson L, Suttle C, Sullivan M. Exploring the Vast Diversity of Marine Viruses. Oceanography. 2007. pp. 135–139. doi:10.5670/oceanog.2007.58

45. Pope WH, Bowman CA, Russell DA, Jacobs-Sera D, Asai DJ, Cresawn SG, *et al.* Whole genome comparison of a large collection of mycobacteriophages reveals a continuum of phage genetic diversity. Elife. 2015;4: e06416.

46. Jordan TC, Burnett SH, Carson S, Caruso SM, Clase K, DeJong RJ, *et al.* A broadly implementable research course in phage discovery and genomics for first-year undergraduate students. MBio. 2014;5: e01051–13.

47. Rohwer F, Edwards R. The Phage Proteomic Tree: a genome-based taxonomy for phage. J Bacteriol. 2002;184: 4529–4535.

48. Mavrich TN, Hatfull GF. Bacteriophage evolution differs by host, lifestyle and genome. Nat Microbiol. 2017;2: 17112.

49. Low SJ, Džunková M, Chaumeil P-A, Parks DH, Hugenholtz P. Evaluation of a concatenated protein phylogeny for classification of tailed double-stranded DNA viruses belonging to the order Caudovirales. Nat Microbiol. 2019;4: 1306–1315.

50. Barylski J, Enault F, Dutilh BE, Schuller MB, Edwards RA, Gillis A, *et al.* Analysis of Spounaviruses as a Case Study for the Overdue Reclassification of

Tailed Phages. Syst Biol. 2020;69: 110–123.

51. Bin Jang H, Bolduc B, Zablocki O, Kuhn JH, Roux S, Adriaenssens EM, *et al.* Taxonomic assignment of uncultivated prokaryotic virus genomes is enabled by gene-sharing networks. Nat Biotechnol. 2019;37: 632–639.

52. Moraru C, Varsani A, Kropinski AM. VIRIDIC-A Novel Tool to Calculate the Intergenomic Similarities of Prokaryote-Infecting Viruses. Viruses. 2020;12. doi:10.3390/v12111268

53. Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics. 2012;28: 3150–3152.

54. Kauffman KM, Chang WK, Brown JM, Hussain FA, Yang J, Polz MF, *et al.* Resolving the structure of phage–bacteria interactions in the context of natural diversity. Nature Communications. 2022. doi:10.1038/s41467-021-27583-z

55. Kupczok A, Neve H, Huang KD, Hoeppner MP, Heller KJ, Franz CMA, *et al.* Rates of Mutation and Recombination in Siphoviridae Phage Genome Evolution over Three Decades. Molecular Biology and Evolution. 2018. pp. 1147–1159. doi:10.1093/molbev/msy027

56. Sanjuán R, Nebot MR, Chirico N, Mansky LM, Belshaw R. Viral Mutation Rates. Journal of Virology. 2010. pp. 9733–9748. doi:10.1128/jvi.00694-10

57. Cuevas JM, Duffy S, Sanjuán R. Point Mutation Rate of Bacteriophage ΦX174. Genetics. 2009. pp. 747–749. doi:10.1534/genetics.109.106005

58. Hynes AP, Rousseau GM, Agudelo D, Goulet A, Amigues B, Loehr J, *et al.* Widespread anti-CRISPR proteins in virulent bacteriophages inhibit a range of Cas9 proteins. Nat Commun. 2018;9: 2919.

59. Domingo-Calap P, Beamud B, Mora-Quilis L, González-Candelas F, Sanjuán R. Isolation and Characterization of Two *Klebsiella pneumoniae* Phages Encoding Divergent Depolymerases. Int J Mol Sci. 2020;21. doi:10.3390/ijms21093160

60. Morris P, Marinelli LJ, Jacobs-Sera D, Hendrix RW, Hatfull GF. Genomic Characterization of Mycobacteriophage Giles: Evidence for Phage Acquisition of Host DNA by Illegitimate Recombination. Journal of Bacteriology. 2008. pp. 2172–2182. doi:10.1128/jb.01657-07

61. Pedulla ML, Ford ME, Houtz JM, Karthikeyan T, Wadsworth C, Lewis JA, *et al.* Origins of highly mosaic mycobacteriophage genomes. Cell. 2003;113: 171–182.

62. Clark AJ, Inwood W, Cloutier T, Dhillon TS. Nucleotide sequence of coliphage HK620 and the evolution of lambdoid phages. Journal of Molecular Biology. 2001. pp. 657–679. doi:10.1006/jmbi.2001.4868

63. De Paepe M, Hutinet G, Son O, Amarir-Bouhram J, Schbath S, Petit M-A. Temperate phages acquire DNA from defective prophages by relaxed homologous recombination: the role of Rad52-like recombinases. PLoS Genet. 2014;10: e1004181.

64. Martinsohn JT, Radman M, Petit M-A. The λ Red Proteins Promote Efficient Recombination between Diverged Sequences: Implications for Bacteriophage Genome Mosaicism. PLoS Genetics. 2008. p. e1000065. doi:10.1371/journal.pgen.1000065

65. Beamud B, Bracho MA, González-Candelas F. Characterization of New Recombinant Forms of HIV-1 From the Comunitat Valenciana (Spain) by Phylogenetic Incongruence. Frontiers in Microbiology. 2019. doi:10.3389/fmicb.2019.01006

66. Pla-Díaz M, Sánchez-Busó L, Giacani L, Šmajs D, Bosshard PP, Bagheri HC, *et al*. Evolutionary Processes in the Emergence and Recent Spread of the Syphilis Agent, *Treponema pallidum*. Mol Biol Evol. 2022;39. doi:10.1093/molbev/msab318

67. Croucher NJ, Page AJ, Connor TR, Delaney AJ, Keane JA, Bentley SD, *et al*. Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. Nucleic Acids Res. 2015;43: e15.

68. Posada D, Crandall KA. The Effect of Recombination on the Accuracy of Phylogeny Estimation. Journal of Molecular Evolution. 2002. pp. 396–402. doi:10.1007/s00239-001-0034-9

69. Sousa JAM de, de Sousa JAM, Pfeifer E, Touchon M, Rocha EPC. Causes and consequences of bacteriophage diversification via genetic exchanges across lifestyles and bacterial taxa. doi:10.1101/2020.04.14.041137

70. Hendrix RW, Lawrence JG, Hatfull GF, Casjens S. The origins and ongoing evolution of viruses. Trends in Microbiology. 2000. pp. 504–508. doi:10.1016/s0966-842x(00)01863-1

71. Hyman P, Abedon ST. Chapter 7 - Bacteriophage Host Range and Bacterial Resistance. Advances in Applied Microbiology. Academic Press; 2010. pp. 217–248.

72. Koskella B, Meaden S. Understanding bacteriophage specificity in

natural microbial communities. Viruses. 2013;5: 806–823.

73. Labrie SJ, Samson JE, Moineau S. Bacteriophage resistance mechanisms. Nat Rev Microbiol. 2010;8: 317–327.

74. Hampton HG, Watson BNJ, Fineran PC. The arms race between bacteria and their phage foes. Nature. 2020;577: 327–336.

75. Mirzaei MK, Nilsson AS. Isolation of Phages for Phage Therapy: A Comparison of Spot Tests and Efficiency of Plating Analyses for Determination of Host Range and Efficacy. PLOS ONE. 2015. p. e0118557. doi:10.1371/journal.pone.0118557

76. Jończyk E, Kłak M, Międzybrodzki R, Górski A. The influence of external factors on bacteriophages--review. Folia Microbiol . 2011;56: 191–200.

77. Abedon ST. Lysis from without. Bacteriophage. 2011;1: 46–49.

78. Rostøl JT, Marraffini L. (Ph)ighting Phages: How Bacteria Resist Their Parasites. Cell Host & Microbe. 2019. pp. 184–194. doi:10.1016/j.chom.2019.01.009

79. Silva JB, Storms Z, Sauvageau D. Host receptors for bacteriophage adsorption. FEMS Microbiology Letters. 2016. p. fnw002. doi:10.1093/femsle/fnw002

80. Zhang Z, Yu F, Zou Y, Qiu Y, Wu A, Jiang T, *et al.* Phage protein receptors have multiple interaction partners and high expressions. Bioinformatics. 2020;36: 2975–2979.

81. Dunne M, Hupfeld M, Klumpp J, Loessner M. Molecular Basis of Bacterial Host Interactions by Gram-Positive Targeting Bacteriophages. Viruses. 2018. p. 397. doi:10.3390/v10080397

82. Maffei E, Shaidullina A, Burkolter M, Heyer Y, Estermann F, Druelle V, *et al.* Systematic exploration of *Escherichia coli* phage–host interactions with the BASEL phage collection. PLOS Biology. 2021. p. e3001424. doi:10.1371/journal.pbio.3001424

83. Dunne M, Prokhorov NS, Loessner MJ, Leiman PG. Reprogramming bacteriophage host range: design principles and strategies for engineering receptor binding proteins. Curr Opin Biotechnol. 2021;68: 272–281.

84. Casjens SR, Molineux IJ. Short noncontractile tail machines: adsorption and DNA delivery by podoviruses. Adv Exp Med Biol. 2012;726: 143–179.

85. Shen Y, Loessner MJ. Beyond antibacterials – exploring bacteriophages as antivirulence agents. Current Opinion in Biotechnology. 2021. pp. 166–173. doi:10.1016/j.copbio.2020.11.004

86. Gordillo Altamirano FL, Barr JJ. Unlocking the next generation of phage therapy: the key is in the receptors. Curr Opin Biotechnol. 2021;68: 115–123.

87. de Jonge PA, Nobrega FL, Brouns SJJ, Dutilh BE. Molecular and Evolutionary Determinants of Bacteriophage Host Range. Trends Microbiol. 2019;27: 51–63.

88. Takeuchi I, Osada K, Azam AH, Asakawa H, Miyanaga K, Tanji Y. The Presence of Two Receptor-Binding Proteins Contributes to the Wide Host Range of Staphylococcal Twort-Like Phages. Applied and Environmental Microbiology. 2016. pp. 5763–5774. doi:10.1128/aem.01385-16

89. Meyer JR, Dobias DT, Weitz JS, Barrick JE, Quick RT, Lenski RE. Repeatability and Contingency in the Evolution of a Key Innovation in Phage Lambda. Science. 2012. pp. 428–432. doi:10.1126/science.1214449

90. Egido JE, Costa AR, Aparicio-Maldonado C, Haas P-J, Brouns SJJ. Mechanisms and clinical importance of bacteriophage resistance. FEMS Microbiol Rev. 2021. doi:10.1093/femsre/fuab048

91. Gordillo Altamirano F, Forsyth JH, Patwa R, Kostoulias X, Trim M, Subedi D, *et al.* Bacteriophage-resistant *Acinetobacter baumannii* are resensitized to antimicrobials. Nat Microbiol. 2021;6: 157–161.

92. Cota I, Sánchez-Romero MA, Hernández SB, Pucciarelli MG, García-Del Portillo F, Casadesús J. Epigenetic Control of *Salmonella enterica* O-Antigen Chain Length: A Tradeoff between Virulence and Bacteriophage Resistance. PLoS Genet. 2015;11: e1005667.

93. Seed KD, Faruque SM, Mekalanos JJ, Calderwood SB, Qadri F, Camilli A. Phase variable O antigen biosynthetic genes control expression of the major protective antigen and bacteriophage receptor in Vibrio cholerae O1. PLoS Pathog. 2012;8: e1002917.

94. Chaudhry W, Lee E, Worthy A, Weiss Z, Grabowicz M, Vega N, *et al.* Mucoidy, a general mechanism for maintaining lytic phage in populations of bacteria. FEMS Microbiol Ecol. 2020;96. doi:10.1093/femsec/fiaa162

95. Cumby N, Edwards AM, Davidson AR, Maxwell KL. The bacteriophage HK97 gp15 moron element encodes a novel superinfection exclusion protein. J Bacteriol. 2012;194: 5012–5019.

96. Johnson AD, Poteete AR, Lauer G, Sauer RT, Ackers GK, Ptashne M. lambda Repressor and cro--components of an efficient molecular switch. Nature. 1981;294: 217–223.

97. Susskind MM, Wright A, Botstein D. Superinfection exclusion by P22 prophage in lysogens of *Salmonella typhimurium*. IV. Genetics and physiology of sieB exclusion. Virology. 1974;62: 367–384.

98. Tesson F, Hervé A, Mordret E, Touchon M, d'Humières C, Cury J, *et al*. Systematic and quantitative view of the antiviral arsenal of prokaryotes. Nat Commun. 2022;13: 2561.

99. Vasu Kommireddy, Nagaraja Valakunja. Diverse Functions of Restriction-Modification Systems in Addition to Cellular Defense. Microbiol Mol Biol Rev. 2013;77: 53–72.

100. Makarova KS, Wolf YI, Alkhnbashi OS, Costa F, Shah SA, Saunders SJ, *et al*. An updated evolutionary classification of CRISPR-Cas systems. Nat Rev Microbiol. 2015;13: 722–736.

101. Makarova KS, Wolf YI, Iranzo J, Shmakov SA, Alkhnbashi OS, Brouns SJJ, *et al*. Evolutionary classification of CRISPR–Cas systems: a burst of class 2 and derived variants. Nature Reviews Microbiology. 2020. pp. 67–83. doi:10.1038/s41579-019-0299-x

102. Barrangou R, Marraffini LA. CRISPR-Cas systems: Prokaryotes upgrade to adaptive immunity. Mol Cell. 2014;54: 234–244.

103. Lopatina A, Tal N, Sorek R. Abortive Infection: Bacterial Suicide as an Antiviral Immune Strategy. Annu Rev Virol. 2020;7: 371–384.

104. Ghafourian S, Raftari M, Sadeghifard N, Sekawi Z. Toxin-antitoxin Systems: Classification, Biological Function and Application in Biotechnology. Curr Issues Mol Biol. 2014;16: 9–14.

105. Penadés JR, Christie GE. The Phage-Inducible Chromosomal Islands: A Family of Highly Evolved Molecular Parasites. Annu Rev Virol. 2015;2: 181–201.

106. Tock MR, Dryden DTF. The biology of restriction and anti-restriction. Current Opinion in Microbiology. 2005. pp. 466–472. doi:10.1016/j.mib.2005.06.003

107. Weigele P, Raleigh EA. Biosynthesis and Function of Modified Bases in Bacteria and Their Viruses. Chem Rev. 2016;116: 12655–12687.

108. Murphy J, Mahony J, Ainsworth S, Nauta A, van Sinderen D. Bacteriophage Orphan DNA Methyltransferases: Insights from Their Bacterial Origin, Function, and Occurrence. Applied and Environmental Microbiology. 2013. pp. 7547–7555. doi:10.1128/aem.02229-13

109. Pawluk A, Davidson AR, Maxwell KL. Anti-CRISPR: discovery, mechanism and function. Nat Rev Microbiol. 2018;16: 12–17.

110. Otsuka Y, Yonesaki T. Dmd of bacteriophage T4 functions as an antitoxin against *Escherichia coli* LsoA and RnlA toxins. Mol Microbiol. 2012;83: 669–681.

111. Isaev A, Drobiazko A, Sierro N, Gordeeva J, Yosef I, Qimron U, *et al*. Phage T7 DNA mimic protein Ocr is a potent inhibitor of BREX defence. Nucleic Acids Research. 2020. doi:10.1093/nar/gkaa510

112. Hussain FA, Dubert J, Elsherbini J, Murphy M, VanInsberghe D, Arevalo P, *et al*. Rapid evolutionary turnover of mobile genetic elements drives bacterial resistance to phages. Science. 2021;374: 488–492.

113. Piel D, Bruto M, Labreuche Y, Blanquart F, Goudenège D, Barcia-Cruz R, *et al*. Phage–host coevolution in natural populations. Nature Microbiology. 2022; 1–12.

114. Janzen DH. When is it Coevolution? Evolution. 1980. p. 611. doi:10.2307/2408229

115. Gandon S, Buckling A, Decaestecker E, Day T. Host-parasite coevolution and patterns of adaptation across time and space. J Evol Biol. 2008;21: 1861–1866.

116. Agrawal, Lively. Infection genetics: gene-for-gene versus matching-alleles models and all points in between. Evol Ecol Res. Available: https://lively.lab.indiana.edu/research/reprints/agrawal_2002_evol-ecol-res.pdf

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

118. Dennehy JJ. What Can Phages Tell Us about Host-Pathogen Coevolution? Int J Evol Biol. 2012;2012: 396165.

119. Flores CO, Meyer JR, Valverde S, Farr L, Weitz JS. Statistical structure of host–phage interactions. Proceedings of the National Academy of Sciences. 2011. doi:10.1073/pnas.1101595108

120. Moebus K, Nattkemper H. Bacteriophage sensitivity patterns among bacteria isolated from marine waters. Helgoländer Meeresuntersuchungen. 1981. pp. 375–385. doi:10.1007/bf02074130

121. Paez-Espino D, Eloe-Fadrosh EA, Pavlopoulos GA, Thomas AD, Huntemann M, Mikhailova N, *et al*. Uncovering Earth's virome. Nature. 2016. pp. 425–430. doi:10.1038/nature19094

122. Wright RCT, Friman V-P, Smith MCM, Brockhurst MA. Cross-resistance is modular in bacteria-phage interactions. PLoS Biol. 2018;16: e2006057.

123. Coclet C, Roux S. Global overview and major challenges of host prediction methods for uncultivated phages. Curr Opin Virol. 2021;49: 117–126.

124. Džunková M, Low SJ, Daly JN, Deng L, Rinke C, Hugenholtz P. Defining the human gut host–phage network through single-cell viral tagging. Nature Microbiology. 2019. pp. 2192–2203. doi:10.1038/s41564-019-0526-2

125. Allers E, Moraru C, Duhaime MB, Beneze E, Solonenko N, Barrero-Canosa J, *et al*. Single-cell and population level viral infection dynamics revealed by phageFISH, a method to visualize intracellular and free viruses. Environ Microbiol. 2013;15: 2306–2318.

126. Labonté JM, Swan BK, Poulos B, Luo H, Koren S, Hallam SJ, *et al*. Single-cell genomics-based analysis of virus-host interactions in marine surface bacterioplankton. ISME J. 2015;9: 2386–2399.

127. Tadmor AD, Ottesen EA, Leadbetter JR, Phillips R. Probing individual environmental bacteria for viruses by using microfluidic digital PCR. Science. 2011;333: 58–62.

128. Marbouty M, Baudry L, Cournac A, Koszul R. Scaffolding bacterial genomes and probing host-virus interactions in gut microbiome by proximity ligation (chromosome capture) assay. Sci Adv. 2017;3: e1602105.

129. de Jonge PA, von Meijenfeldt FAB, Costa AR, Nobrega FL, Brouns SJJ, Dutilh BE. Adsorption Sequencing as a Rapid Method to Link Environmental Bacteriophages to Hosts. iScience. 2020;23: 101439.

130. Edwards RA, McNair K, Faust K, Raes J, Dutilh BE. Computational approaches to predict bacteriophage-host relationships. FEMS Microbiol Rev. 2016;40: 258–272.

131. Khan Mirzaei M, Deng L. New technologies for developing phage-

based tools to manipulate the human microbiome. Trends Microbiol. 2022;30: 131–142.

132. Dutilh BE, Cassman N, McNair K, Sanchez SE, Silva GGZ, Boling L, *et al*. A highly abundant bacteriophage discovered in the unknown sequences of human faecal metagenomes. Nat Commun. 2014;5: 4498.

133. Bobay L-M, Touchon M, Rocha EPC. Pervasive domestication of defective prophages by bacteria. Proc Natl Acad Sci U S A. 2014;111: 12127–12132.

134. Westra ER, Brouns SJJ. The rise and fall of CRISPRs--dynamics of spacer acquisition and loss. Molecular microbiology. 2012. pp. 1021–1025.

135. Nami Y, Imeni N, Panahi B. Application of machine learning in bacteriophage research. BMC Microbiol. 2021;21: 193.

136. Lood C, Boeckaerts D, Stock M, De Baets B, Lavigne R, van Noort V, *et al*. Digital phagograms: predicting phage infectivity through a multilayer machine learning approach. Curr Opin Virol. 2022;52: 174–181.

137. Boeckaerts D, Stock M, Criel B, Gerstmans H, De Baets B, Briers Y. Predicting bacteriophage hosts based on sequences of annotated receptor-binding proteins. Sci Rep. 2021;11: 1467.

138. Dion MB, Plante P-L, Zufferey E, Shah SA, Corbeil J, Moineau S. Streamlining CRISPR spacer-based bacterial host predictions to decipher the viral dark matter. Nucleic Acids Res. 2021;49: 3127–3138.

139. Young F, Rogers S, Robertson DL. Predicting host taxonomic information from viral genomes: A comparison of feature representations. PLoS Comput Biol. 2020;16: e1007894.

140. Villarroel J, Kleinheinz KA, Jurtz VI, Zschach H, Lund O, Nielsen M, *et al*. HostPhinder: A Phage Host Prediction Tool. Viruses. 2016;8. doi:10.3390/v8050116

141. Coutinho FH, Zaragoza-Solas A, López-Pérez M, Barylski J, Zielezinski A, Dutilh BE, *et al*. RaFAH: Host prediction for viruses of Bacteria and Archaea based on protein content. Patterns (N Y). 2021;2: 100274.

142. Lamy-Besnier Q, Brancotte B, Ménager H, Debarbieux L. Viral Host Range database, an online tool for recording, analyzing and disseminating virus–host interactions. Bioinformatics. 2021. pp. 2798–2801. doi:10.1093/bioinformatics/btab070

143. Hryckowian AJ, Merrill BD, Porter NT, Van Treuren W, Nelson EJ, Garlena RA, *et al*. Bacteroides thetaiotaomicron-Infecting Bacteriophage Isolates Inform Sequence-Based Host Range Predictions. Cell Host Microbe. 2020;28: 371–379.e5.

144. Whitfield C, Wear SS, Sande C. Assembly of Bacterial Capsular Polysaccharides and Exopolysaccharides. Annu Rev Microbiol. 2020;74: 521–543.

145. Candela T, Fouet A. Poly-gamma-glutamate in bacteria. Molecular Microbiology. 2006. pp. 1091–1098. doi:10.1111/j.1365-2958.2006.05179.x

146. Rendueles O, Garcia-Garcerà M, Néron 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.

147. Nucci A, Rocha E, Rendueles O. Adaptation to novel spatially-structured environments is driven by the capsule and alters virulence-associated traits. Research Square. 2022. doi:10.21203/rs.3.rs-1320833/v1

148. 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.

149. Haudiquet M, Buffet A, Rendueles O, Rocha EPC. Interplay between the cell envelope and mobile genetic elements shapes gene flow in populations of the nosocomial pathogen Klebsiella pneumoniae. PLoS Biol. 2021;19: e3001276.

150. Zhensong W, Zhang J-R. Bacterial Capsules. Molecular Medical Microbiology. unknown; 2015. pp. 33–53.

151. Campos MA, Vargas MA, Regueiro V, Llompart CM, Albertí S, Bengoechea JA. Capsule polysaccharide mediates bacterial resistance to antimicrobial peptides. Infect Immun. 2004;72: 7107–7114.

152. Lawlor MS, Hsu J, Rick PD, Miller VL. Identification of Klebsiella pneumoniae virulence determinants using an intranasal infection model. Mol Microbiol. 2005;58: 1054–1073.

153. Weintraub A. Immunology of bacterial polysaccharide antigens. Carbohydr Res. 2003;338: 2539–2547.

154. Ernst CM, Braxton JR, Rodriguez-Osorio CA, Zagieboylo AP, Li L, Pironti A, *et al*. Adaptive evolution of virulence and persistence in carbapenem-

resistant Klebsiella pneumoniae. Nat Med. 2020;26: 705–711.

155. Buffet A, Rocha EPC, Rendueles O. Nutrient conditions are primary drivers of bacterial capsule maintenance in. Proc Biol Sci. 2021;288: 20202876.

156. Lam MMC, Wick RR, Judd LM, Holt KE, Wyres KL. Kaptive 2.0: updated capsule and lipopolysaccharide locus typing for the Klebsiella pneumoniae species complex. Microbial Genomics. 2022;8: 000800.

157. Holt KE, Lassalle F, Wyres KL, Wick R, Mostowy RJ. Diversity and evolution of surface polysaccharide synthesis loci in Enterobacteriales. ISME J. 2020;14: 1713–1730.

158. Nanayakkara BS, O'Brien CL, Gordon DM. Diversity and distribution of Klebsiella capsules in Escherichia coli. Environ Microbiol Rep. 2019;11: 107–117.

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

160. Wagner A. Robustness and evolvability: a paradox resolved. Proc Biol Sci. 2008;275: 91–100.

161. Adams MH, Park BH. An enzyme produced by a phage-host cell system. II. The properties of the polysaccharide depolymerase. Virology. 1956;2: 719–736.

162. Pires DP, Oliveira H, Melo LDR, Sillankorva S, Azeredo J. Bacteriophage-encoded depolymerases: their diversity and biotechnological applications. Appl Microbiol Biotechnol. 2016;100: 2141–2151.

163. Latka A, Leiman PG, Drulis-Kawa Z, Briers Y. Modeling the Architecture of Depolymerase-Containing Receptor Binding Proteins in Klebsiella Phages. Front Microbiol. 2019;10: 2649.

164. Knecht LE, Veljkovic M, Fieseler L. Diversity and Function of Phage Encoded Depolymerases. Front Microbiol. 2019;10: 2949.

165. Leiman PG, Battisti AJ, Bowman VD, Stummeyer K, Mühlenhoff M, Gerardy-Schahn R, *et al*. The structures of bacteriophages K1E and K1-5 explain processive degradation of polysaccharide capsules and evolution of new host specificities. J Mol Biol. 2007;371: 836–849.

166. Pieroni P, Rennie RP, Ziola B, Deneer HG. The use of bacteriophages to differentiate serologically cross-reactive isolates of Klebsiella pneumoniae. J Med

Microbiol. 1994;41: 423–429.

167. Smug B, Szczepaniak K, Rocha EPC, Dunin-Horkawicz S, Mostowy RJ. Protein modularity in phages is extensive and associated with functions linked to core replication machinery and host tropism determinants. bioRxiv. 2022. p. 2022.12.27.521992. doi:10.1101/2022.12.27.521992

168. Oliveira H, Costa AR, Konstantinides N, Ferreira A, Akturk E, Sillankorva S, *et al*. Ability of phages to infect *Acinetobacter calcoaceticus-Acinetobacter baumannii* complex species through acquisition of different pectate lyase depolymerase domains. Environmental Microbiology. 2021. pp. 3334–3334. doi:10.1111/1462-2920.15620

169. Roach DR, Sjaarda DR, Castle AJ, Svircev AM. Host exopolysaccharide quantity and composition impact Erwinia amylovora bacteriophage pathogenesis. Appl Environ Microbiol. 2013;79: 3249–3256.

170. Ouyang R, Costa AR, Cassidy CK, Otwinowska A, Williams VCJ, Latka A, *et al*. High resolution reconstruction of a Jumbo bacteriophage infecting capsulated bacteria using hyperbranched tail fibers. bioRxiv. 2022. doi:10.1101/2022.03.15.484430

171. Pan Y-J, Lin T-L, Chen C-C, Tsai Y-T, Cheng Y-H, Chen Y-Y, *et al*. Klebsiella Phage ΦK64-1 Encodes Multiple Depolymerases for Multiple Host Capsular Types. J Virol. 2017;91. doi:10.1128/JVI.02457-16

172. Latka A, Lemire S, Grimon D, Dams D, Maciejewska B, Lu T, *et al*. Engineering the Modular Receptor-Binding Proteins of Klebsiella Phages Switches Their Capsule Serotype Specificity. mBio. 2021. doi:10.1128/mbio.00455-21

173. de Sousa JAM, Buffet A, Haudiquet M, Rocha EPC, Rendueles O. Modular prophage interactions driven by capsule serotype select for capsule loss under phage predation. ISME J. 2020;14: 2980–2996.

174. Prokhorov NS, Riccio C, Zdorovenko EL, Shneider MM, Browning C, Knirel YA, *et al*. Function of bacteriophage G7C esterase tailspike in host cell adsorption. Mol Microbiol. 2017;105: 385–398.

175. Born Y, Fieseler L, Klumpp J, Eugster MR, Zurfluh K, Duffy B, *et al*. The tail-associated depolymerase of Erwinia amylovora phage L1 mediates host cell adsorption and enzymatic capsule removal, which can enhance infection by other phage. Environ Microbiol. 2014;16: 2168–2180.

176. Hesse S, Rajaure M, Wall E, Johnson J, Bliskovsky V, Gottesman S, *et*

al. Phage Resistance in Multidrug-Resistant Klebsiella pneumoniae ST258 Evolves via Diverse Mutations That Culminate in Impaired Adsorption. MBio. 2020;11. doi:10.1128/mBio.02530-19

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

178. Majkowska-Skrobek G, Markwitz P, Sosnowska E, Lood C, Lavigne R, Drulis-Kawa Z. The evolutionary trade-offs in phage-resistant Klebsiella pneumoniae entail cross-phage sensitization and loss of multidrug resistance. Environ Microbiol. 2021. doi:10.1111/1462-2920.15476

179. Gladstone EG, Molineux IJ, Bull JJ. Evolutionary principles and synthetic biology: avoiding a molecular tragedy of the commons with an engineered phage. J Biol Eng. 2012;6: 13.

180. Oliveira H, Mendes A, Fraga AG, Ferreira A, Pimenta AI, Mil-Homens D, *et al*. K2 Capsule Depolymerase Is Highly Stable, Is Refractory to Resistance, and Protects Larvae and Mice from Acinetobacter baumannii Sepsis. Appl Environ Microbiol. 2019;85. doi:10.1128/AEM.00934-19

181. Pan Y-J, Lin T-L, Chen Y-Y, Lai P-H, Tsai Y-T, Hsu C-R, *et al*. Identification of three podoviruses infecting Klebsiella encoding capsule depolymerases that digest specific capsular types. Microb Biotechnol. 2019;12: 472–486.

182. Fang Q, Feng Y, McNally A, Zong Z. Characterization of phage resistance and phages capable of intestinal decolonization of carbapenem-resistant Klebsiella pneumoniae in mice. Commun Biol. 2022;5: 48.

183. Rendueles O, de Sousa JAM, Rocha EPC. Competition between phage-resistance mechanisms determines the outcome of bacterial co-existence. bioRxiv. 2022. p. 2022.07.11.499539. doi:10.1101/2022.07.11.499539

184. Porter NT, Hryckowian AJ, Merrill BD, Fuentes JJ, Gardner JO, Glowacki RWP, *et al*. Phase-variable capsular polysaccharides and lipoproteins modify bacteriophage susceptibility in Bacteroides thetaiotaomicron. Nat Microbiol. 2020;5: 1170–1181.

185. Pelkonen S, Aalto J, Finne J. Differential activities of bacteriophage depolymerase on bacterial polysaccharide: binding is essential but degradation is inhibitory in phage infection of K1-defective Escherichia coli. J Bacteriol. 1992;174: 7757–7761.

186. Chapman-McQuiston E, Wu XL. Stochastic receptor expression allows

sensitive bacteria to evade phage attack. Part I: experiments. Biophys J. 2008;94: 4525–4536.

187. Aidley J, Sørensen MCH, Bayliss CD, Brøndsted L. Phage exposure causes dynamic shifts in the expression states of specific phase-variable genes of Campylobacter jejuni. Microbiology. 2017;163: 911–919.

188. Markwitz P, Lood C, Olszak T, van Noort V, Lavigne R, Drulis-Kawa Z. Genome-driven elucidation of phage-host interplay and impact of phage resistance evolution on bacterial fitness. ISME J. 2022;16: 533–542.

189. Stummeyer K, Schwarzer D, Claus H, Vogel U, Gerardy-Schahn R, Mühlenhoff M. Evolution of bacteriophages infecting encapsulated bacteria: lessons from Escherichia coli K1-specific phages. Mol Microbiol. 2006;60: 1123–1135.

190. Chattaway MA, Schaefer U, Tewolde R, Dallman TJ, Jenkins C. Identification of Escherichia coli and Shigella Species from Whole-Genome Sequences. Journal of Clinical Microbiology. 2017. pp. 616–623. doi:10.1128/jcm.01790-16

191. Wyres KL, Lam MMC, Holt KE. Population genomics of Klebsiella pneumoniae. Nat Rev Microbiol. 2020;18: 344–359.

192. Patro LPP, Sudhakar KU, Rathinavelan T. K-PAM: a unified platform to distinguish Klebsiella species K- and O-antigen types, model antigen structures and identify hypervirulent strains. Sci Rep. 2020;10: 16732.

193. Venturini C, Ben Zakour NL, Bowring B, Morales S, Cole R, Kovach Z, *et al*. Fine capsule variation affects bacteriophage susceptibility in Klebsiella pneumoniae ST258. FASEB J. 2020;34: 10801–10817.

194. 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 U S A. 2015;112: E3574–81.

195. Hennart M, Guglielmini J, Bridel S, Maiden MCJ, Jolley KA, Criscuolo A, *et al*. A Dual Barcoding Approach to Bacterial Strain Nomenclature: Genomic Taxonomy of Klebsiella pneumoniae Strains. Mol Biol Evol. 2022;39. doi:10.1093/molbev/msac135

196. Diancourt L, Passet V, Verhoef J, Grimont PAD, Brisse S. Multilocus sequence typing of Klebsiella pneumoniae nosocomial isolates. J Clin Microbiol. 2005;43: 4178–4182.

197. Jolley KA, Bray JE, Maiden MCJ. Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications. Wellcome Open Res. 2018;3: 124.

198. 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.

199. Hawkey J, Wyres KL, Judd LM, Harshegyi T, Blakeway L, Wick RR, *et al*. ESBL plasmids in Klebsiella pneumoniae: diversity, transmission and contribution to infection burden in the hospital setting. Genome Med. 2022;14: 97.

200. Ngiam L, Weynberg KD, Guo J. The presence of plasmids in bacterial hosts alters phage isolation and infectivity. ISME Communications. 2022;2. doi:10.1038/s43705-022-00158-9

201. Gorrie CL, Mirceta M, Wick RR, Edwards DJ, Thomson NR, Strugnell RA, *et al*. Gastrointestinal Carriage Is a Major Reservoir of Klebsiella pneumoniae Infection in Intensive Care Patients. Clin Infect Dis. 2017;65: 208–215.

202. Wyres KL, . Wyres KL, Holt KE. Klebsiella pneumoniae as a key trafficker of drug resistance genes from environmental to clinically important bacteria. Current Opinion in Microbiology. 2018. pp. 131–139. doi:10.1016/j.mib.2018.04.004

203. Davis GS, Price LB. Recent Research Examining Links Among Klebsiella pneumoniae from Food, Food Animals, and Human Extraintestinal Infections. Current Environmental Health Reports. 2016. pp. 128–135. doi:10.1007/s40572-016-0089-9

204. Herridge WP, Shibu P, O'Shea J, Brook TC, Hoyles L. Bacteriophages of spp., their diversity and potential therapeutic uses. J Med Microbiol. 2020;69: 176–194.

205. Domingo-Calap P, Beamud B, Vienne J, González-Candelas F, Sanjuán R. Isolation of Four Lytic Phages Infecting K22 Clinical Isolates from Spain. Int J Mol Sci. 2020;21. doi:10.3390/ijms21020425

206. Townsend EM, Kelly L, Gannon L, Muscatt G, Dunstan R, Michniewski S, *et al*. Isolation and Characterization of Phages for Phage Therapy. Phage (New Rochelle). 2021;2: 26–42.

207. Lourenço M, Osbelt L, Passet V, Gravey F, Strowig T, Rodrigues C, *et al*. Phages against non-capsulated *Klebsiella pneumoniae*: broader host range, slower resistance. bioRxiv. 2022. doi:10.1101/2022.08.04.502604

208. García-González N, Beamud B, Fuster B, Giner S, Domínguez MV, Sánchez A, *et al.* Tracking the Emergence and Dissemination of a *bla* $_{NDM-23}$ Gene in a Multidrug Resistance Plasmid of Klebsiella pneumoniae. Microbiology Spectrum. 2023. doi:10.1128/spectrum.02585-22

209. Djukovic A, González-Barberá EM, Sanz J, Artacho A, Peñaranda I, Herrera B, *et al.* High Heterogeneity of Multidrug-Resistant *Enterobacteriaceae* Fecal Levels in Hospitalized Patients Is Partially Driven by Intravenous β-Lactams. Antimicrobial Agents and Chemotherapy. 2020. doi:10.1128/aac.01415-19

210. Wu K-M, Li L-H, Yan J-J, Tsao N, Liao T-L, Tsai H-C, et al. Genome sequencing and comparative analysis of Klebsiella pneumoniae NTUH-K2044, a strain causing liver abscess and meningitis. J Bacteriol. 2009;191: 4492–4501.

211. Brisse S, Passet V, Haugaard AB, Babosan A, Kassis-Chikhani N, Struve C, *et al.* wzi Gene sequencing, a rapid method for determination of capsular type for Klebsiella strains. J Clin Microbiol. 2013;51: 4073–4078.

212. Larsson A. AliView: a fast and lightweight alignment viewer and editor for large datasets. Bioinformatics. 2014. pp. 3276–3278. doi:10.1093/bioinformatics/btu531

213. Chao K-H, Barton K, Palmer S, Lanfear R. sangeranalyseR: Simple and Interactive Processing of Sanger Sequencing Data in R. Genome Biol Evol. 2021;13. doi:10.1093/gbe/evab028

214. Staden R. The Staden sequence analysis package. Mol Biotechnol. 1996;5: 233–241.

215. Domenico P, Schwartz S, Cunha BA. Reduction of capsular polysaccharide production in Klebsiella pneumoniae by sodium salicylate. Infect Immun. 1989;57: 3778–3782.

216. Blumenkrantz N, Asboe-Hansen G. New method for quantitative determination of uronic acids. Anal Biochem. 1973;54: 484–489.

217. Luria SE, Delbruck M. Mutations of Bacteria From Virus Sensitivity to Virus Resistance. 1943.

218. Prjibelski A, Antipov D, Meleshko D, Lapidus A, Korobeynikov A. Using SPAdes De Novo Assembler. Curr Protoc Bioinformatics. 2020;70: e102.

219. Wick RR, Judd LM, Gorrie CL, Holt KE. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. PLOS Computational Biology. 2017. p. e1005595. doi:10.1371/journal.pcbi.1005595

220. Schmieder R, Edwards R. Quality control and preprocessing of metagenomic datasets. Bioinformatics. 2011;27: 863–864.

221. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics. 2014;30: 2068–2069.

222. Lam MMC, Wick RR, Watts SC, Cerdeira LT, Wyres KL, Holt KE. A genomic surveillance framework and genotyping tool for Klebsiella pneumoniae and its related species complex. Nat Commun. 2021;12: 4188.

223. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics. 2009;25: 1972–1973.

224. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, *et al*. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. Mol Biol Evol. 2020;37: 1530–1534.

225. Felsenstein J. Phylogenies From Restriction Sites: A Maximum-likelihood Approach. Evolution. 1992. pp. 159–173. doi:10.1111/j.1558-5646.1992.tb01991.x

226. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods. 2017;14: 587–589.

227. Kans J. Entrez Direct: E-utilities on the Unix Command Line. Entrez Programming Utilities Help [Internet]. National Center for Biotechnology Information (US); 2022.

228. UniProt Consortium. UniProt: the universal protein knowledgebase in 2021. Nucleic Acids Res. 2021;49: D480–D489.

229. Bayliss SC, Thorpe HA, Coyle NM, Sheppard SK, Feil EJ. PIRATE: A fast and scalable pangenomics toolbox for clustering diverged orthologues in bacteria. Gigascience. 2019;8. doi:10.1093/gigascience/giz119

230. Sirén K, Millard A, Petersen B, Gilbert MTP, Clokie MRJ, Sicheritz-Pontén T. Rapid discovery of novel prophages using biological feature engineering and machine learning. doi:10.1101/2020.08.09.243022

231. Russel J, Pinilla-Redondo R, Mayo-Muñoz D, Shah SA, Sørensen SJ. CRISPRCasTyper: Automated Identification, Annotation, and Classification of CRISPR-Cas Loci. The CRISPR Journal. 2020. pp. 462–469. doi:10.1089/crispr.2020.0059

232. Millman A, Melamed S, Amitai G, Sorek R. Diversity and classification of cyclic-oligonucleotide-based anti-phage signalling systems. Nat Microbiol. 2020;5: 1608–1615.

233. Eddy SR. Accelerated Profile HMM Searches. PLoS Comput Biol. 2011;7: e1002195.

234. Abby SS, Néron B, Ménager H, Touchon M, Rocha EPC. MacSyFinder: a program to mine genomes for molecular systems with an application to CRISPR-Cas systems. PLoS One. 2014;9: e110726.

235. Kauffman KM, Polz MF. Streamlining standard bacteriophage methods for higher throughput. MethodsX. 2018;5: 159–172.

236. Jakočiūnė D, Moodley A. A Rapid Bacteriophage DNA Extraction Method. Methods Protoc. 2018;1. doi:10.3390/mps1030027

237. Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. Nat Methods. 2012;9: 671–675.

238. Garneau JR, Depardieu F, Fortier L-C, Bikard D, Monot M. PhageTerm: a tool for fast and accurate determination of phage termini and packaging mechanism using next-generation sequencing data. Sci Rep. 2017;7: 8292.

239. Website. Available: Andrews, S. (2010). FastQC: A Quality Control Tool for High Throughput Sequence Data [Online]. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

240. Ewels P, Magnusson M, Lundin S, Käller M. MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics. 2016. pp. 3047–3048. doi:10.1093/bioinformatics/btw354

241. Wood DE, Salzberg SL. Kraken: ultrafast metagenomic sequence classification using exact alignments. Genome Biol. 2014;15: R46.

242. Breitwieser FP, Salzberg SL. Pavian: Interactive analysis of metagenomics data for microbiomics and pathogen identification. doi:10.1101/084715

243. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, *et al.* SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 2012;19: 455–477.

244. Russell DA. Sequencing, Assembling, and Finishing Complete Bacteriophage Genomes. Methods Mol Biol. 2018;1681: 109–125.

245. Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. Nat Commun. 2018;9: 5114.

246. Paradis E, Schliep K. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics. 2019;35: 526–528.

247. Carroll LM, Wiedmann M, Kovac J. Proposal of a Taxonomic Nomenclature for the Bacillus cereus Group Which Reconciles Genomic Definitions of Bacterial Species with Clinical and Industrial Phenotypes. MBio. 2020;11. doi:10.1128/mBio.00034-20

248. Nishimura Y, Yoshida T, Kuronishi M, Uehara H, Ogata H, Goto S. ViPTree: the viral proteomic tree server. Bioinformatics. 2017;33: 2379–2380.

249. Letunic I, Bork P. Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy. Nucleic Acids Res. 2011;39: W475–8.

250. Terzian P, Ndela EO, Galiez C, Lossouarn J, Bucio REP, Mom R, *et al*. PHROG: families of prokaryotic virus proteins clustered using remote homology. NAR Genomics and Bioinformatics. 2021. doi:10.1093/nargab/lqab067

251. Shen W, Le S, Li Y, Hu F. SeqKit: A Cross-Platform and Ultrafast Toolkit for FASTA/Q File Manipulation. PLoS One. 2016;11: e0163962.

252. Guy L, Roat Kultima J, Andersson SGE. genoPlotR: comparative gene and genome visualization in R. Bioinformatics. 2010. pp. 2334–2335. doi:10.1093/bioinformatics/btq413

253. Tynecki P, Guziński A, Kazimierczak J, Jadczuk M, Dastych J, Onisko A. PhageAI - Bacteriophage Life Cycle Recognition with Machine Learning and Natural Language Processing. Cold Spring Harbor Laboratory. 2020. p. 2020.07.11.198606. doi:10.1101/2020.07.11.198606

254. McNair K, Bailey BA, Edwards RA. PHACTS, a computational approach to classifying the lifestyle of phages. Bioinformatics. 2012;28: 614–618.

255. Hockenberry AJ, Wilke CO. BACPHLIP: Predicting bacteriophage lifestyle from conserved protein domains. doi:10.1101/2020.05.13.094805

256. Luo E, Aylward FO, Mende DR, DeLong EF. Bacteriophage Distributions and Temporal Variability in the Ocean's Interior. MBio. 2017;8. doi:10.1128/mBio.01903-17

257. Wetzel KS, Aull HG, Zack KM, Garlena RA, Hatfull GF. Protein-

Mediated and RNA-Based Origins of Replication of Extrachromosomal Mycobacterial Prophages. MBio. 2020;11. doi:10.1128/mBio.00385-20

258. Doster E, Lakin SM, Dean CJ, Wolfe C, Young JG, Boucher C, *et al*. MEGARes 2.0: a database for classification of antimicrobial drug, biocide and metal resistance determinants in metagenomic sequence data. Nucleic Acids Res. 2020;48: D561–D569.

259. Chen L, Zheng D, Liu B, Yang J, Jin Q. VFDB 2016: hierarchical and refined dataset for big data analysis--10 years on. Nucleic Acids Res. 2016;44: D694–7.

260. Boeckaerts D, Stock M, De Baets B, Briers Y. Identification of Phage Receptor-Binding Protein Sequences with Hidden Markov Models and an Extreme Gradient Boosting Classifier. Viruses. 2022;14. doi:10.3390/v14061329

261. Steinegger M, Meier M, Mirdita M, Vöhringer H, Haunsberger SJ, Söding J. HH-suite3 for fast remote homology detection and deep protein annotation. BMC Bioinformatics. 2019;20: 473.

262. Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, *et al*. InterProScan 5: genome-scale protein function classification. Bioinformatics. 2014;30: 1236–1240.

263. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE. The Phyre2 web portal for protein modeling, prediction and analysis. Nature Protocols. 2015. pp. 845–858. doi:10.1038/nprot.2015.053

264. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, *et al*. Highly accurate protein structure prediction with AlphaFold. Nature. 2021;596: 583–589.

265. van Kempen M, Kim S, Tumescheit C, Mirdita M, Gilchrist CLM, Soeding J, *et al*. Foldseek: fast and accurate protein structure search. bioRxiv. 2022. doi:10.1101/2022.02.07.479398

266. Castresana J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Biol Evol. 2000;17: 540–552.

267. Garnier J, Gibrat JF, Robson B. GOR method for predicting protein secondary structure from amino acid sequence. Methods Enzymol. 1996;266: 540–553.

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

1028.

269. Rajnovic D, Muñoz-Berbel X, Mas J. Fast phage detection and quantification: An optical density-based approach. PLoS One. 2019;14: e0216292.

270. Beal J, Farny NG, Haddock-Angelli T, Selvarajah V, Baldwin GS, Buckley-Taylor R, *et al.* Robust estimation of bacterial cell count from optical density. Commun Biol. 2020;3: 512.

271. Xie Y, Wahab L, Gill JJ. Development and Validation of a Microtiter Plate-Based Assay for Determination of Bacteriophage Host Range and Virulence. Viruses. 2018;10. doi:10.3390/v10040189

272. Wright RCT, Friman V-P, Smith MCM, Brockhurst MA. Resistance Evolution against Phage Combinations Depends on the Timing and Order of Exposure. MBio. 2019;10. doi:10.1128/mBio.01652-19

273. Website. Available: Andri et mult. al. S (2021). DescTools: Tools for Descriptive Statistics. R package version 0.99.44, https://cran.r-project.org/package=DescTools.

274. Aphalo PJ. Learn R. 2020. doi:10.1201/9780429060342

275. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez J-C, *et al.* pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics. 2011;12: 77.

276. Ratan A. Assembly Algorithms for Next-generation Sequence Data. 2009.

277. Wyres KL, Holt KE. Klebsiella pneumoniae Population Genomics and Antimicrobial-Resistant Clones. Trends Microbiol. 2016;24: 944–956.

278. Li H-Y, Kao C-Y, Lin W-H, Zheng P-X, Yan J-J, Wang M-C, *et al.* Characterization of CRISPR-Cas Systems in Clinical Klebsiella pneumoniae Isolates Uncovers Its Potential Association With Antibiotic Susceptibility. Frontiers in Microbiology. 2018. doi:10.3389/fmicb.2018.01595

279. Wang G, Song G, Xu Y. Association of CRISPR/Cas System with the Drug Resistance in Klebsiella pneumoniae. Infection and Drug Resistance. 2020. pp. 1929–1935. doi:10.2147/idr.s253380

280. Roach MJ, McNair K, Giles SK, Inglis L, Pargin E, Roux S, *et al.* Philympics 2021: Prophage Predictions Perplex Programs.

doi:10.1101/2021.06.03.446868

281. Barondess JJ, Beckwith J. bor gene of phage lambda, involved in serum resistance, encodes a widely conserved outer membrane lipoprotein. J Bacteriol. 1995;177: 1247–1253.

282. Swanson NA, Cingolani G. A Tail of Phage Adhesins. Structure . 2018. pp. 1565–1567.

283. Bonilla E, Costa AR, van den Berg DF, van Rossum T, Hagedoorn S, Walinga H, *et al.* Genomic characterization of four novel bacteriophages infecting the clinical pathogen Klebsiella pneumoniae. DNA Res. 2021;28. doi:10.1093/dnares/dsab013

284. Crépin T, Swale C, Monod A, Garzoni F, Chaillet M, Berger I. Polyproteins in structural biology. Current Opinion in Structural Biology. 2015. pp. 139–146. doi:10.1016/j.sbi.2015.04.007

285. 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: 3103–3119.

286. Daubie V, Chalhoub H, Blasdel B, Dahma H, Merabishvili M, Glonti T, et al. Determination of phage susceptibility as a clinical diagnostic tool: A routine perspective. Front Cell Infect Microbiol. 2022;12: 1000721.

287. Chiarelli A, Cabanel N, Rosinski-Chupin I, Zongo PD, Naas T, Bonnin RA, et al. Diversity of mucoid to non-mucoid switch among carbapenemase-producing Klebsiella pneumoniae. BMC Microbiol. 2020;20: 325.

288. Zhang Y, Skolnick J. TM-align: a protein structure alignment algorithm based on the TM-score. Nucleic Acids Res. 2005;33: 2302–2309.

289. Bielen A, Ćetković H, Long PF, Schwab H, Abramić M, Vujaklija D. The SGNH-hydrolase of Streptomyces coelicolor has (aryl)esterase and a true lipase activity. Biochimie. 2009. pp. 390–400. doi:10.1016/j.biochi.2008.10.018

290. Ho TD, Slauch JM. OmpC is the receptor for Gifsy-1 and Gifsy-2 bacteriophages of Salmonella. J Bacteriol. 2001;183: 1495–1498.

291. Kobe B, Kajava AV. The leucine-rich repeat as a protein recognition motif. Curr Opin Struct Biol. 2001;11: 725–732.

292. Dunstan RA, Bamert RS, Belousoff MJ, Short FL, Barlow CK, Pickard

DJ, *et al*. Mechanistic Insights into the Capsule-Targeting Depolymerase from a Klebsiella pneumoniae Bacteriophage. Microbiol Spectr. 2021;9: e0102321.

293. Sørensen PE, Ng DYK, Duchateau L, Ingmer H, Garmyn A, Butaye P. Classification of In Vitro Phage–Host Population Growth Dynamics. Microorganisms. 2021;9: 2470.

294. Xie Y, Wahab L, Gill JJ. Development and Validation of a Microtiter Plate-Based Assay for Determination of Bacteriophage Host Range and Virulence. Viruses. 2018;10: 189.

295. Yehl K, Lemire S, Yang AC, Ando H, Mimee M, Torres MDT, *et al*. Engineering Phage Host-Range and Suppressing Bacterial Resistance through Phage Tail Fiber Mutagenesis. Cell. 2019;179: 459–469.e9.

296. Abedon ST, Yin J. Bacteriophage plaques: theory and analysis. Methods Mol Biol. 2009;501: 161–174.

297. Abedon ST. The Murky Origin of Snow White and Her T-Even Dwarfs. Genetics. 2000;155: 481–486.

298. Casjens SR, Hendrix RW. Bacteriophage lambda: Early pioneer and still relevant. Virology. 2015. pp. 310–330. doi:10.1016/j.virol.2015.02.010

299. Olsen NS, Forero-Junco L, Kot W, Hansen LH. Exploring the Remarkable Diversity of Culturable Escherichia coli Phages in the Danish Wastewater Environment. Viruses. 2020. p. 986. doi:10.3390/v12090986

300. Kauffman KM, Brown JM, Sharma RS, VanInsberghe D, Elsherbini J, Polz M, *et al*. Viruses of the Nahant Collection, characterization of 251 marine Vibrionaceae viruses. Scientific Data. 2018. doi:10.1038/sdata.2018.114

301. Göller PC, Elsener T, Lorgé D, Radulovic N, Bernardi V, Naumann A, *et al*. Multi-species host range of staphylococcal phages isolated from wastewater. Nat Commun. 2021;12: 6965.

302. Huang X, Li X, An H, Wang J, Ding M, Wang L, *et al*. Capsule type defines the capability of Klebsiella pneumoniae in evading Kupffer cell capture in the liver. PLoS Pathog. 2022;18: e1010693.

303. Rocha J, Henriques I, Gomila M, Manaia CM. Common and distinctive genomic features of Klebsiella pneumoniae thriving in the natural environment or in clinical settings. Scientific Reports. 2022. doi:10.1038/s41598-022-14547-6

304. Bleriot I, Blasco L, Pacios O, Fernández-García L, López M, Ortiz-

Cartagena C, *et al*. Molecular analysis of the interactions between phages and the bacterial host *Klebsiella pneumoniae*. doi:10.1101/2022.09.12.507515

305. Jurczak-Kurek A, Gąsior T, Nejman-Faleńczyk B, Bloch S, Dydecka A, Topka G, et al. Biodiversity of bacteriophages: morphological and biological properties of a large group of phages isolated from urban sewage. Sci Rep. 2016;6: 34338.

306. 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: 378.

307. Fong K, Tremblay DM, Delaquis P, Goodridge L, Levesque RC, Moineau S, *et al*. Diversity and Host Specificity Revealed by Biological Characterization and Whole Genome Sequencing of Bacteriophages Infecting. Viruses. 2019;11. doi:10.3390/v11090854

308. Flores CO, Valverde S, Weitz JS. Multi-scale structure and geographic drivers of cross-infection within marine bacteria and phages. ISME J. 2013;7: 520–532.

309. Emslander Q, Vogele K, Braun P, Stender J, Willy C, Joppich M, *et al*. Cell-free production of personalized therapeutic phages targeting multidrug-resistant bacteria. Cell Chem Biol. 2022;29: 1434–1445.e7.

310. Epping L, van Tonder AJ, Gladstone RA, The Global Pneumococcal Sequencing Consortium, Bentley SD, Page AJ, *et al*. SeroBA: rapid high-throughput serotyping of Streptococcus pneumoniae from whole genome sequence data. Microb Genom. 2018;4. doi:10.1099/mgen.0.000186

311. Watts SC, Holt KE. hicap: Serotyping of the Haemophilus influenzae Capsule Locus. J Clin Microbiol. 2019;57. doi:10.1128/JCM.00190-19

312. Bessonov K, Laing C, Robertson J, Yong I, Ziebell K, Gannon VPJ, *et al*. ECTyper: in silico Escherichia coli serotype and species prediction from raw and assembled whole-genome sequence data. Microbial Genomics. 2021. doi:10.1099/mgen.0.000728

313. Adler BA, Kazakov AE, Zhong C, Liu H, Kutter E, Lui LM, *et al*. The genetic basis of phage susceptibility, cross-resistance and host-range in Salmonella. Microbiology. 2021;167. doi:10.1099/mic.0.001126

314. Berrington AW, Tan Y-C, Srikhanta Y, Kuipers B, van der Ley P, Peak IRA, *et al*. Phase variation in meningococcal lipooligosaccharide biosynthesis genes. FEMS Immunol Med Microbiol. 2002;34: 267–275.

315. Li J, Zhang J-R. Phase Variation of Streptococcus pneumoniae. Gram-Positive Pathogens, Third Edition. American Society of Microbiology; 2019. pp. 331–343.

316. Taketani M, Donia MS, Jacobson AN, Lambris JD, Fischbach MA. A Phase-Variable Surface Layer from the Gut Symbiont Bacteroides thetaiotaomicron. MBio. 2015;6: e01339–15.

317. Villa L, Feudi C, Fortini D, Brisse S, Passet V, Bonura C, *et al*. Diversity, virulence, and antimicrobial resistance of the KPC-producing ST307 clone. Microb Genom. 2017;3: e000110.

318. Scholl D, Adhya S, Merril CR. Bacteriophage SP6 is closely related to phages K1-5, K5, and K1E but encodes a tail protein very similar to that of the distantly related P22. J Bacteriol. 2002;184: 2833–2836.

319. Gencay YE, Gambino M, Prüssing TF, Brøndsted L. The genera of bacteriophages and their receptors are the major determinants of host range. Environ Microbiol. 2019;21: 2095–2111.

320. Sørensen AN, Woudstra C, Holst Sørensen MC, Brøndsted L. Subtypes of tail spike proteins predicts the host range of Ackermannviridae phages. Computational and Structural Biotechnology Journal. 2021. pp. 4854–4867. doi:10.1016/j.csbj.2021.08.030

321. Haines MEK, Hodges FE, Nale JY, Mahony J, van Sinderen D, Kaczorowska J, *et al.* Analysis of selection methods to develop novel phage therapy cocktails against antimicrobial resistant clinical isolates of bacteria. Cold Spring Harbor Laboratory. 2020. p. 2020.10.13.337345. doi:10.1101/2020.10.13.337345

322. Maffei E, Burkolter M, Heyer Y, Egli A, Jenal U, Harms A. Phage Paride hijacks bacterial stress responses to kill dormant, antibiotic-tolerant cells. bioRxiv. 2022. doi:10.1101/2022.01.26.477855

323. Favre-Bonte S, Joly B, Forestier C. Consequences of reduction of Klebsiella pneumoniae capsule expression on interactions of this bacterium with epithelial cells. Infect Immun. 1999;67: 554–561.

324. Costa AR, van den Berg DF, Esser JQ, Muralidharan A, van den Bossche H, Bonilla BE, *et al.* Accumulation of defense systems drives panphage resistance in *Pseudomonas aeruginosa*. bioRxiv. 2022. doi:10.1101/2022.08.12.503731

325. Snyder L. Phage-exclusion enzymes: a bonanza of biochemical and cell biology reagents? Molecular Microbiology. 1995. pp. 415–420.

doi:10.1111/j.1365-2958.1995.tb02255.x

326. Cain AK, Barquist L, Goodman AL, Paulsen IT, Parkhill J, van Opijnen T. A decade of advances in transposon-insertion sequencing. Nat Rev Genet. 2020;21: 526–540.

327. Rousset F, Cui L, Siouve E, Becavin C, Depardieu F, Bikard D. Genome-wide CRISPR-dCas9 screens in E. coli identify essential genes and phage host factors. PLOS Genetics. 2018. p. e1007749. doi:10.1371/journal.pgen.1007749

328. 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: 2854–2866.

329. Johnston CD, Cotton SL, Rittling SR, Starr JR, Borisy GG, Dewhirst FE, *et al*. Systematic evasion of the restriction-modification barrier in bacteria. Proc Natl Acad Sci U S A. 2019;116: 11454–11459.

330. Furuta Y, Namba-Fukuyo H, Shibata TF, Nishiyama T, Shigenobu S, Suzuki Y, *et al*. Methylome diversification through changes in DNA methyltransferase sequence specificity. PLoS Genet. 2014;10: e1004272.

331. Pfeifer E, Moura de Sousa JA, Touchon M, Rocha EPC. Bacteria have numerous distinctive groups of phage-plasmids with conserved phage and variable plasmid gene repertoires. Nucleic Acids Res. 2021;49: 2655–2673.

332. Tan D, Zhang Y, Qin J, Le S, Gu J, Chen L-K, *et al*. A Frameshift Mutation in wcaJ Associated with Phage Resistance in Klebsiella pneumoniae. Microorganisms. 2020;8: 378.

333. Castledine M, Padfield D, Sierocinski P, Pascual JS, Hughes A, Mäkinen L, *et al*. Parallel evolution of Pseudomonas aeruginosa phage resistance and virulence loss in response to phage treatment in vivo and in vitro. eLife. 2022. doi:10.7554/elife.73679

334. Bertozzi Silva J, Storms Z, Sauvageau D. Host receptors for bacteriophage adsorption. FEMS Microbiol Lett. 2016;363. doi:10.1093/femsle/fnw002

335. Tétart F, Repoila F, Monod C, Krisch HM. Bacteriophage T4 host range is expanded by duplications of a small domain of the tail fiber adhesin. J Mol Biol. 1996;258: 726–731.

336. Lees-Miller RG, Iwashkiw JA, Scott NE, Seper A, Vinogradov E, Schild S, *et al*. A common pathway for O-linked protein-glycosylation and synthesis of

capsule in Acinetobacter baumannii. Mol Microbiol. 2013;89: 816–830.

337. Sellner B, Prakapaitė R, van Berkum M, Heinemann M, Harms A, Jenal U. A New Sugar for an Old Phage: a c-di-GMP-Dependent Polysaccharide Pathway Sensitizes Escherichia coli for Bacteriophage Infection. MBio. 2021;12: e0324621.

338. Ren G, Wang Z, Li Y, Hu X, Wang X. Effects of Lipopolysaccharide Core Sugar Deficiency on Colanic Acid Biosynthesis in Escherichia coli. J Bacteriol. 2016;198: 1576–1584.

339. Izquierdo L, Coderch N, Piqué N, Bedini E, Corsaro MM, Merino S, *et al*. The Klebsiella pneumoniae wabG gene: role in biosynthesis of the core lipopolysaccharide and virulence. J Bacteriol. 2003;185: 7213–7221.

340. Cai R, Wang G, Le S, Wu M, Cheng M, Guo Z, *et al*. Three Capsular Polysaccharide Synthesis-Related Glucosyltransferases, GT-1, GT-2 and WcaJ, Are Associated With Virulence and Phage Sensitivity of Klebsiella pneumoniae. Frontiers in Microbiology. 2019. doi:10.3389/fmicb.2019.01189

341. Singh S, Wilksch JJ, Dunstan RA, Mularski A, Wang N, Hocking D, *et al*. LPS O Antigen Plays a Key Role in Klebsiella pneumoniae Capsule Retention. Microbiol Spectr. 2022;10: e0151721.

342. Lee H, Shin J, Chung Y-J, Baek JY, Chung DR, Peck KR, *et al*. Evolution of Klebsiella pneumoniae with mucoid and non-mucoid type colonies within a single patient. Int J Med Microbiol. 2019;309: 194–198.

343. Zhao S, Lieberman TD, Poyet M, Kauffman KM, Gibbons SM, Groussin M, *et al*. Adaptive Evolution within Gut Microbiomes of Healthy People. Cell Host Microbe. 2019;25: 656–667.e8.

344. Dapa T, Ramiro RS, Pedro MF, Gordo I, Xavier KB. Diet leaves a genetic signature in a keystone member of the gut microbiota. Cell Host Microbe. 2022;30: 183–199.e10.

345. Hayes S, Vincentelli R, Mahony J, Nauta A, Ramond L, Lugli GA, *et al*. Functional carbohydrate binding modules identified in evolved dits from siphophages infecting various Gram-positive bacteria. Mol Microbiol. 2018;110: 777–795.

346. Lavelle K, Goulet A, McDonnell B, Spinelli S, van Sinderen D, Mahony J, *et al*. Revisiting the host adhesion determinants of Streptococcus thermophilus siphophages. Microb Biotechnol. 2020;13: 1765–1779.

347. Born F, Braun P, Scholz HC, Grass G. Specific Detection of Yersinia pestis Based on Receptor Binding Proteins of Phages. Pathogens. 2020. p. 611. doi:10.3390/pathogens9080611

348. Hsu C-R, Liao C-H, Lin T-L, Yang H-R, Yang F-L, Hsieh P-F, *et al*. Identification of a capsular variant and characterization of capsular acetylation in Klebsiella pneumoniae PLA-associated type K57. Sci Rep. 2016;6: 31946.

349. Connolly JPR, Turner NCA, Serrano E, Rimbi PT, Browning DF, O'Boyle N, *et al*. Control of resistance against bacteriophage killing by a metabolic regulator in meningitis-associated *Escherichia coli*. Proc Natl Acad Sci U S A. 2022;119: e2210299119.

350. Deszo EL, Steenbergen SM, Freedberg DI, Vimr ER. Escherichia coli K1 polysialic acid O-acetyltransferase gene, neuO, and the mechanism of capsule form variation involving a mobile contingency locus. Proc Natl Acad Sci U S A. 2005;102. doi:10.1073/pnas.0407428102

351. 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: 5635–5646.

352. Bono LM, Gensel CL, Pfennig DW, Burch CL. Evolutionary rescue and the coexistence of generalist and specialist competitors: an experimental test. Proc Biol Sci. 2015;282: 20151932.

353. Latino L, Pourcel C. Recovery and Characterization of Bacteria Resisting Infection by Lytic Bacteriophage. Methods Mol Biol. 2018;1693: 85–98.

354. Burmeister AR, Fortier A, Roush C, Lessing AJ, Bender RG, Barahman R, *et al*. Pleiotropy complicates a trade-off between phage resistance and antibiotic resistance. Proc Natl Acad Sci U S A. 2020;117: 11207–11216.

355. Knopp M, Andersson DI. Amelioration of the Fitness Costs of Antibiotic Resistance Due To Reduced Outer Membrane Permeability by Upregulation of Alternative Porins. Mol Biol Evol. 2015;32: 3252–3263.

356. Koskella B, Brockhurst MA. Bacteria-phage coevolution as a driver of ecological and evolutionary processes in microbial communities. FEMS Microbiol Rev. 2014;38: 916–931.

357. Millman A, Melamed S, Leavitt A, Doron S, Bernheim A, Hör J, *et al*. An expanded arsenal of immune systems that protect bacteria from phages. Cell Host Microbe. 2022. doi:10.1016/j.chom.2022.09.017

358. Wu Y, van den Hurk A, Aparicio-Maldonado C, Kushwaha SK, King CM, Ou Y, *et al*. Defence systems provide synergistic anti-phage activity in E. coli. bioRxiv. 2022. p. 2022.08.21.504612. doi:10.1101/2022.08.21.504612

359. Lenski RE, Levin BR. Constraints on the Coevolution of Bacteria and Virulent Phage: A Model, Some Experiments, and Predictions for Natural Communities. Am Nat. 1985;125: 585–602.

360. Lourenço M, Chaffringeon L, Lamy-Besnier Q, Pédron T, Campagne P, Eberl C, *et al*. The Spatial Heterogeneity of the Gut Limits Predation and Fosters Coexistence of Bacteria and Bacteriophages. Cell Host Microbe. 2020;28: 390–401.e5.

361. Schrag SJ, Mittler JE. Host-Parasite Coexistence: The Role of Spatial Refuges in Stabilizing Bacteria-Phage Interactions. The American Naturalist. 1996. pp. 348–377. doi:10.1086/285929

362. Strathdee SA, Hatfull GF, Mutalik VK, Schooley RT. Phage therapy: From biological mechanisms to future directions. Cell. 2023. pp. 17–31. doi:10.1016/j.cell.2022.11.017

363. Pelfrene E, Sebris Z, Cavaleri M. Comment on Fauconnier, A. Phage Therapy Regulation: From Night to Dawn. Viruses. 2019;11. doi:10.3390/v11090771

364. McCallin S, Sacher JC, Zheng J, Chan BK. Current State of Compassionate Phage Therapy. Viruses. 2019;11. doi:10.3390/v11040343

365. Uyttebroek S, Chen B, Onsea J, Ruythooren F, Debaveye Y, Devolder D, *et al*. Safety and efficacy of phage therapy in difficult-to-treat infections: a systematic review. Lancet Infect Dis. 2022;22: e208–e220.

366. Antimicrobial Resistance Collaborators. Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. Lancet. 2022;399: 629–655.

367. Pirnay J-P, De Vos D, Verbeken G, Merabishvili M, Chanishvili N, Vaneechoutte M, *et al*. The phage therapy paradigm: prêt-à-porter or sur-mesure? Pharm Res. 2011;28: 934–937.

368. Nair A, Khairnar K. Genetically engineered phages for therapeutics: proceed with caution. Nat Med. 2019;25: 1028.

369. Altamirano FLG, Gordillo Altamirano FL, Barr JJ. Phage Therapy in the Postantibiotic Era. Clinical Microbiology Reviews. 2019. doi:10.1128/cmr.00066-

370. 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. Frontiers in Microbiology. 2018. doi:10.3389/fmicb.2018.02517

371. Lin H, Paff ML, Molineux IJ, Bull JJ. Therapeutic Application of Phage Capsule Depolymerases against K1, K5, and K30 Capsulated E. coli in Mice. Frontiers in Microbiology. 2017. doi:10.3389/fmicb.2017.02257

372. Chan BK, Abedon ST, Loc-Carrillo C. Phage cocktails and the future of phage therapy. Future Microbiol. 2013;8: 769–783.

373. Dorman MJ, Feltwell T, Goulding DA, Parkhill J, Short FL. The Capsule Regulatory Network of *Klebsiella pneumoniae* Defined by density-TraDISort. mBio. 2018. doi:10.1128/mbio.01863-18

374. Camarillo-Guerrero LF, Almeida A, Rangel-Pineros G, Finn RD, Lawley TD. Massive expansion of human gut bacteriophage diversity. Cell. 2021;184: 1098–1109.e9.

375. Peters SL, Borges AL, Giannone RJ, Morowitz MJ, Banfield JF, Hettich RL. Experimental validation that human microbiome phages use alternative genetic coding. Nature Communications. 2022. doi:10.1038/s41467-022-32979-6

376. Yang W, Lin Y-C, Johnson W, Dai N, Vaisvila R, Weigele P, *et al*. A genome-phenome association study in native microbiomes identifies a mechanism for cytosine modification in DNA and RNA. Elife. 2021;10. doi:10.7554/eLife.70021

377. Neri U, Wolf YI, Roux S, Camargo AP, Lee B, Kazlauskas D, *et al*. Expansion of the global RNA virome reveals diverse clades of bacteriophages. Cell. 2022;185: 4023–4037.e18.

378. Wohlfarth JC, Feldmüller M, Schneller A, Kilcher S, Burkolter M, Meile S, *et al*. L-form conversion in Gram-positive bacteria enables escape from phage infection. Nature Microbiology. 2023. doi:10.1038/s41564-022-01317-3

379. Hardy A, Kever L, Frunzke J. Antiphage small molecules produced by bacteria – beyond protein-mediated defenses. Trends in Microbiology. 2023. pp. 92–106. doi:10.1016/j.tim.2022.08.001

380. Rousset F, Yirmiya E, Nesher S, Brandis A, Mehlman T, Itkin M, *et al*. A

conserved family of immune effectors cleaves cellular ATP upon viral infection. doi:10.1101/2023.01.24.525353

381. Cury J, Mordret E, Trejo VH, Tesson F, Ofir G, Poirier EZ, et al. Conservation of antiviral systems across domains of life reveals novel immune mechanisms in humans. doi:10.1101/2022.12.12.520048

382. Ibarra-Chávez R, Hansen MF, Pinilla-Redondo R, Seed KD, Trivedi U. Phage satellites and their emerging applications in biotechnology. FEMS Microbiol Rev. 2021;45. doi:10.1093/femsre/fuab031

383. Eppley JM, Biller SJ, Luo E, Burger A, DeLong EF. Marine viral particles reveal an expansive repertoire of phage-parasitizing mobile elements. Proc Natl Acad Sci U S A. 2022;119: e2212722119.

384. Payne LJ, Meaden S, Mestre MR, Palmer C, Toro N, Fineran PC, *et al*. PADLOC: a web server for the identification of antiviral defence systems in microbial genomes. Nucleic Acids Res. 2022;50: W541–50.

385. Federici S, Kredo-Russo S, Valdés-Mas R, Kviatcovsky D, Weinstock E, Matiuhin Y, *et al*. Targeted suppression of human IBD-associated gut microbiota commensals by phage consortia for treatment of intestinal inflammation. Cell. 2022;185: 2879–2898.e24.

386. Nogueira CL, Pires DP, Monteiro R, Santos SB, Carvalho CM. Exploitation of a Bacteriophage Receptor-Binding Protein as a Superior Biorecognition Molecule. ACS Infect Dis. 2021;7: 3077–3087.

8. Supplementary material

All the supplementary material of this work is available in the Figshare repository of this thesis. A digital object identifier (doi) and the link are included for each item. Alternatively, all supplementary material can be downloaded in one Excel file (10.6084/m9.figshare.22259959).

Supplementary materials include the following:

Table S1. Review of *Klebsiella* phages from the literature. 10.6084/m9.figshare.22259812.

Table S2. General features of the *Klebsiella* clinical strains used in this study. The strain, sequencing details, virulence and antibiotic resistance scores, *wzi* allele, CLT and OLT inference, number of prophages, CRISPR loci, and presence of defense systems and secondary receptors are provided. Isolation hosts of phages are also indicated. The presence of a capsule was determined by visual inspection of colonies using light microscopy. 10.6084/m9.figshare.22259827.

Table S3. Isolation and sequencing results of the phages included in this study. Isolation details and characteristics of isolation hosts are provided along with sequencing results. Assembly strategy: 1. Trimmed reads assembled with Unicycler. 2. Trimmed reads assembled with SPAdes. 3. Raw reads assembled with SPAdes; 4. Subset of raw reads assembled with MIRA4 using k-mer sizes of 149, 171 and 193. 10.6084/m9.figshare.22259833.

Table S4. List of proteins and protein names used to detect phage receptors. 10.6084/m9.figshare.22259836.

Table S5. List of proteins with depolymerase activity from the literature. **"**Experimental" indicates whether proteins were recombinantly expressed and their activity against the capsule validated. These proteins were used to detect depolymerases in phage genomes using blast. 10.6084/m9.figshare.22259839.

Table S6. Bacterial genome sequences from the NLSAR and Infect-ERA consortium analysed in this work. From this dataset representative strains were selected to build the Kpn collection. 10.6084/m9.figshare.22259845.

Table S7. Main features of the 46 *Klebsiella* phage representatives. Phage ID, phylogenetic classification in 13 groups, genome size, GC %, CDS, putative lysogeny genes and lifestyle prediction are provided. The taxonomic classification according to IGS values with the closest phage and the updated ICTV taxonomy (accessed 17/02/23) are included. *: Taxonomy ranks marked with a star can be reclassified based on the relationships shown in this study; **: Taxonomic ranks pending formal acceptance. Additionally, plaque morphologies and sizes, concentration of phages for the spot assay, susceptible CLTs, RBP system and number of depolymerase domains (Dpos) are also shown. CLTs marked with a star indicate that bacteria were acapsular. Dpos marked with a star indicate putative soluble Dpos. REP: transcription repressor/regulator, PAR: parAB genes, REC: recombinases. TSP: Tail spike protein. CBM: carbohydrate-binding module. 10.6084/m9.figshare.22259851.

Table S8. Phage-bacteria positive combinations. The strain used for phage isolation is indicated. Spotting: positive by the spot-assay. Virulent: positive by the killing planktonic or progeny assay. 10.6084/m9.figshare.22259857.

Table S9. Results of depolymerization assays. H: presence of plaque haloes. Gluc200: µg of uronic acid per 200 µL of capsule suspension. The mean from two biological replicates is shown**.** Control refers to bacteria treated with SM buffer instead with phage. 10.6084/m9.figshare.22259860.

Table S10. Clustering of phage RBDs with putative depolymerase activity from this work and the literature. RBDs are identified by the phage name followed by the phage protein in which the domain was detected, and the positions considered after removal of the anchor domain. 10.6084/m9.figshare.22259863.

Table S11. Measurements of phage depolymerization in avirulent combinations of bacteria and phage (Dpo.activity%). P-values were obtained comparing the µg of uronic acid per 200 µL of capsule suspension in phage-treated and untreated bacteria by the t.test (sd = standard deviation). 10.6084/m9.figshare.22259869.

Table S12. Measurements of phage adsorption in avirulent combinations of bacteria and phage. Results for reversible adsorption are shown (Ads.rate; sd =

standard deviation). Results for irreversible adsorption did not differ from these. 10.6084/m9.figshare.22259878.

Table S13. RBPs of S8/S9 phages and associated domains found with InterProScan. 10.6084/m9.figshare.22259881

Table S14. Quantification of Cap- mutants in bacterial cultures of the strain collection. Acap-p1 indicates the number of Cap- mutants that remained acapsular after one passage. Note that this information is only available for a subset of all Cap- mutants. 10.6084/m9.figshare.22259884.

Table S15. Infectivity of Cap- mutants by phages by the bactericidal assay. 10.6084/m9.figshare.22259887.

Table S16. Analysis of RBPs of phages without Dpo domains. The first hit with an annotated and relevant function from Foldseek is shown. TM represents the template modeling score for the query and target protein structures, with values closer to one indicating higher similarity. 10.6084/m9.figshare.22259890.

Table S17. Quantification of phage-bacteria dynamics from bacterial growth curves. 10.6084/m9.figshare.22259893.