Genomic and Phenotypic Characterisation of Bacteriophages for the Biocontrol of *Klebsiella* spp.

Claire Karen Anita Elek

March 2024

86,442 words

A thesis submitted for the degree of Doctor of Philosophy

Quadram Institute Bioscience

University of East Anglia

ABSTRACT

Klebsiella spp. represent a major threat to human health due to their ability to cause a wide range of infections, particularly in healthcare settings, and the prevalence of antimicrobial resistance determinants amongst them. The use of bacteriophages (phages) as an alternative or adjuvant to antibiotics has recently seen a resurgence and this thesis aimed to investigate the use of phages against *Klebsiella*. This project identified 26 genetically distinct phages infecting a range of well-characterised *Klebsiella* spp., with phage isolates spanning five known genera and two new genera within four families in the class *Caudoviricetes*. Phage time-kill assays revealed two- and three-phage combinations were more efficacious at killing and suppressing bacterial regrowth, particularly when the phages were from distinct genetic lineages. Synergy was observed between phage and antibiotics but was highly dependent on the phage-antibiotic combination. The implementation and optimisation of phage amplification and purification protocols revealed challenges in generating high-quality, high-titre phage stocks. Phage therapy also requires complete, accurate, high-quality phage genomes, and previous work has found traditional phage assembly methods to be error-prone. The work here describes the development of a new workflow: HYPPA – a HYbrid and Poly-polish Phage Assembly to improve phage genomics. This methodology utilises long-read assemblies in combination with short-read sequencing for resolving phage genomes and correcting assembly errors, negating the need for extensive manual curation. Lastly, a transposon mutant library was generated in a *Klebsiella michiganensis* strain to investigate phage resistance and susceptibility genes in a functional genomics screen. This work provides new insights into understanding the phenotype-genotype link through functional assays and genomics, and the important role complete, accurate, high-quality phage genomes play for regulatory approval of phages as therapeutics in the UK.

Table of Contents

LIST OF FIGURES

head is full so that each linear piece of DNA contains more than 100% of the genome (bottom panel). The DNA is packaged into the capsid, which will then contain multiple copies of the genome with different starting sequences. Figure generated in BioRender and adapted from Moore and Prevelige (2002); Merrill *et al.* (2016); Hilbert *et al.* (2017). 22

LIST OF APPENDICES

LIST OF ABBREVIATIONS AND ACRONYMS

AAHC	Antibiotic-associated haemorrhagic colitis	FIC	Fractional inhibitory concentration
AME	Aminoglycoside-modifying enzymes	FPLC	Fast purification liquid chromatography
AMR	Antimicrobial resistance	*g*	Gravity at sea level
ANI	Average nucleotide identity	GC	Guanine-cytosine
BBSRC	Biotechnology and Biological Sciences Research Council	GMP	Good Manufacturing Practice
		h	Hour(s)
BHI	Brain-heat infusion	HAI	Healthcare-associated infection
BLAST	Basic local alignment search tool	HAP	Healthcare-associated pneumonia
bp	Base pairs	HGT	Horizontal gene transfer
CAR	Carbenicillin	HPC	High performance computer
CAUTI	Catheter-associated urinary tract infection	hpi	Hours post-inoculation
CDS	Coding sequence	HvKP	Hypervirulent *K. pneumoniae*
CFU	Colony forming unit		
CG	Clonal group	IBD	Inflammatory bowel disease
CHL	Chloramphenicol	ICTV	International Committee on Taxonomy of Viruses
CIP	Ciprofloxacin		
CPS	Capsular polysaccharide	Indel(s)	Insertions/deletion(s)
CRKP	Carbapenem-resistant *K. pneumoniae*	IRIDA	Integrated Rapid Infectious Disease Analysis
CST	Colistin		
CTX	Cefotaxime	KAN	Kanamycin
DEAE	Diethylethanolamine	kbp	Kilobase pairs
DNA	Deoxyribonucleic acid	kDa	Kilodalton
ds	Double-stranded	KL	Capsular antigen locus
DTR	Direct terminal repeat	LAL	Limulus amebocyte lysate
EDTA	Ethylenediaminetetraacetic acid	LB	Lysogeny broth
		LOD	Limit of detection
EOP	Efficiency of plaquing	LPS	Lipopolysaccharide
ESBL	Extended-spectrum β-lactamase	M	Molar
EU	Endotoxin unit	MDR	Multidrug resistant
EUCAST	the European Committee on Antimicrobial Susceptibility Testing	MEM	Meropenem
		MHA	Muellar-hinton agar
		MHB	Muellar-hinton broth

MIC	Minimum inhibitory concentration		RNAP	RNA polymerase
MIM	Minimum inhibitory MOI		r.p.m.	Rotations per minute
min	Minute(s)		SCV	Single colony variant
ML	Maximum-likelihood		SD	Standard deviation
MLST	Multilocus sequence typing		SDS	Sodium dodecyl sulfate
MOI	Multiplicity of inoculation		SDS-PAGE	SDS-polyacrylamide gel electrophoresis
mRNA	Messenger RNA		SNP	Single-nucleotide polymorphism
MSA	Multisequence alignment		ss	Single-stranded
NCBI	National Centre for Biotechnology Information		ST	Sequence type
NEB	New England Biolabs		STR	Streptomycin
NEC	Necrotising enterocolitis		TEM	Transmission electron microscopy
NGS	Next-generation sequencing		TerL	Terminase large subunit
NHS	National Health Service		TerS	Terminase small subunit
NICE	National Institute for Health and Care Excellence		TET	Tetracycline
			TFF	Tangential flow filtration
OD	Optical density		TIS	Transposon-insertion sequencing
OL	O-antigen locus		TMP	Trimethoprim
OM	Outer membrane		TraDIS	Transposon-directed insertion site sequencing
ONT	Oxford Nanopore Technologies		tRNA	Transfer RNA
PCR	Polymerase chain reaction		TZP	Piperacillin/tazobactam
PDR	Pan-drug resistant		UIS	Unique insertion site
PEG	Polyethylene glycol		UK	United Kingdom
PIP	Piperacillin		USA	United States of America
PFU	Plaque forming unit		UTI	Urinary tract infection
ppm	Parts per million		UV	Ultraviolet
PVDF	Polyvinylidene fluoride		VAP	Ventilator-associated pneumonia
QA	Quaternary amine		VFDB	Virulence factor database
QC	Quality control		WGS	Whole-genome sequencing
QIB	Quadram Institute Bioscience		WHO	World Health Organization
RBP	Receptor-binding protein		WT	Wild-type
RI	Restriction inhibitor		XDR	Extensively-drug resistant
RM	Restriction-modification			
RNA	Ribonucleic acid			

ACKNOWLEDGEMENTS

I would first like to express my deepest thanks and gratitude to my supervisor Evelien Adriaenssens. Without her incredible patience, support, and belief in my ability I do not think I would have been able to complete this PhD. It takes a unique individual to be as adaptable and as encouraging as she has been with me over the last 4+ years. I would also like to thank my secondary supervisor Mark Webber whose support and encouragement has also contributed considerably to my PhD experience in such a positive way. I also want to thank Teagan Brown and Hannah Pye for guidance and support, both in and out of the lab.

I have been very privileged to work with so many incredible scientists over the course of my PhD including all members, past, present, and honorary, of the Adriaenssens Group. I am very grateful for all their contributions, whether that be professional or personal. I'd like to thank both Emma and Yasir for their help with all the TraDIS work and for allowing me to bombard them both with questions. I'd like to thank everyone at the Quadram Institute that has contributed so much to my PhD, including all the bioinformaticians, the sequencing team members, and members of the imaging teams across the Norwich Research Park. I would also like to thank all collaborators from the UK and abroad who provided assistance and guidance.

I'd like to thank my family for their love, support, as well as the sacrifices they have made for me to achieve my PhD. To my parents who taught me that hard work pays off and to my siblings Lauren and Luke for always believing in me. To my daughters Ava and Adeleine, who make me proud every day and who have known little outside of me in education. To my husband and partner-in-wine David, whose patience, kindness, and love has gotten me through many, many dark times. I am in awe of him every day, and I feel so privileged to be his wife. He is the Tokugawa to my Oda.

Lastly, I would like to dedicate this thesis to people I have lost on the way. To Dr Oliver Charity, whose kindness, brilliance, and passion for science shines on through us all. To my little-big brother Marky aka Whistle, who was so talented and bright, and whose absence is felt so deeply still.

This thesis is for anyone that struggles or has struggled with their mental health. You are enough.

CHAPTER 1: INTRODUCTION

1.1 *Klebsiella* species

The genus *Klebsiella* comprises a heterogenous group of Gram-negative bacteria in the family *Enterobacteriaceae* (Podschun and Ullman, 1998). The members are encapsulated, non-sporulating, nonmotile (except for *Klebsiella aerogenes*) (Merla *et al.*, 2019), and are often asymptomatically associated with human and animal mucosae (Holt *et al.*, 2015). *Klebsiella* spp. are environmentally ubiquitous, having been isolated from soil, surface water, food, and sewage (Podschun and Ullman, 1998; Holt *et al.*, 2015; Paczosa and Mecsas, 2016; Merla *et al.*, 2019; Crippa *et al.*, 2023).

1.1.1 *Klebsiella* diversity

Klebsiella pneumoniae and *Klebsiella oxytoca* are considered the most clinically relevant of the *Klebsiella* spp., but extensive genetic diversity within the genus has resulted in the adoption of the terms *K. pneumoniae* species complex and *K. oxytoca* species complex to describe their population structure (**Figure 1.1**) (Podschun and Ullman, 1998; Brisse and Verhoef, 2001; Holt *et al.*, 2015). The *K. pneumoniae* and *K. oxytoca* complexes both comprise several different species, subspecies, and phylogroups (**Table 1.1**). Within the *K. pneumoniae* species complex are *K. pneumoniae*, of which there are three subspecies (*sensu stricto*); *K. quasipneumoniae*, of which there are two subspecies; *K. variicola*, of which there are two subspecies; *K. quasivariicola*, and *K. africana*. Within the *K. oxytoca* species complex are *K. oxytoca* (*sensu stricto*); *K. grimontii*; *K. huaxiensis*; *K. michiganensis*; *K. pasteurii*; and *K. spallanzanii*.

The complex population structure of *Klebsiella* spp. complicates identification in clinical microbiology laboratories, as they do not differentiate the *Klebsiella* spp. within their complexes (Fontana *et al.*, 2019). This can lead to an underestimation of these lesser-known species and subspecies in both asymptomatic carriage and infections (Maatallah *et al.*, 2014; Fontana *et al.*, 2019). Further complicating the identification of *Klebsiella* spp. are the overlapping biochemical characteristics and molecular dissimilarities, whereby whole genome sequencing (WGS) is the only reliable method of determining speciation (Maatallah *et al.*, 2014; Fontana *et al.*, 2019).

Advancements in genomic analyses have led to a growing number of available *Klebsiella* genomes in databases and the increasing reclassification of many of these genomes. Several studies have reported misclassification of *K. variicola* infections as *K. pneumoniae* (Seki *et al.*, 2013; Maatallah *et al.*, 2014; Berry *et al.*, 2015; Long *et al.*, 2017; Fontana *et al.*, 2019). Moreover, *K. grimontii* isolates recovered from several infection sites were originally misclassified as *K. oxytoca* (Herzog *et al.*, 2014; Passet and Brisse, 2018), and WGS identified *K. michiganensis* as the causative organism in a bacteraemia case that was initially assigned *K. oxytoca* (Seiffert *et al.*, 2019).
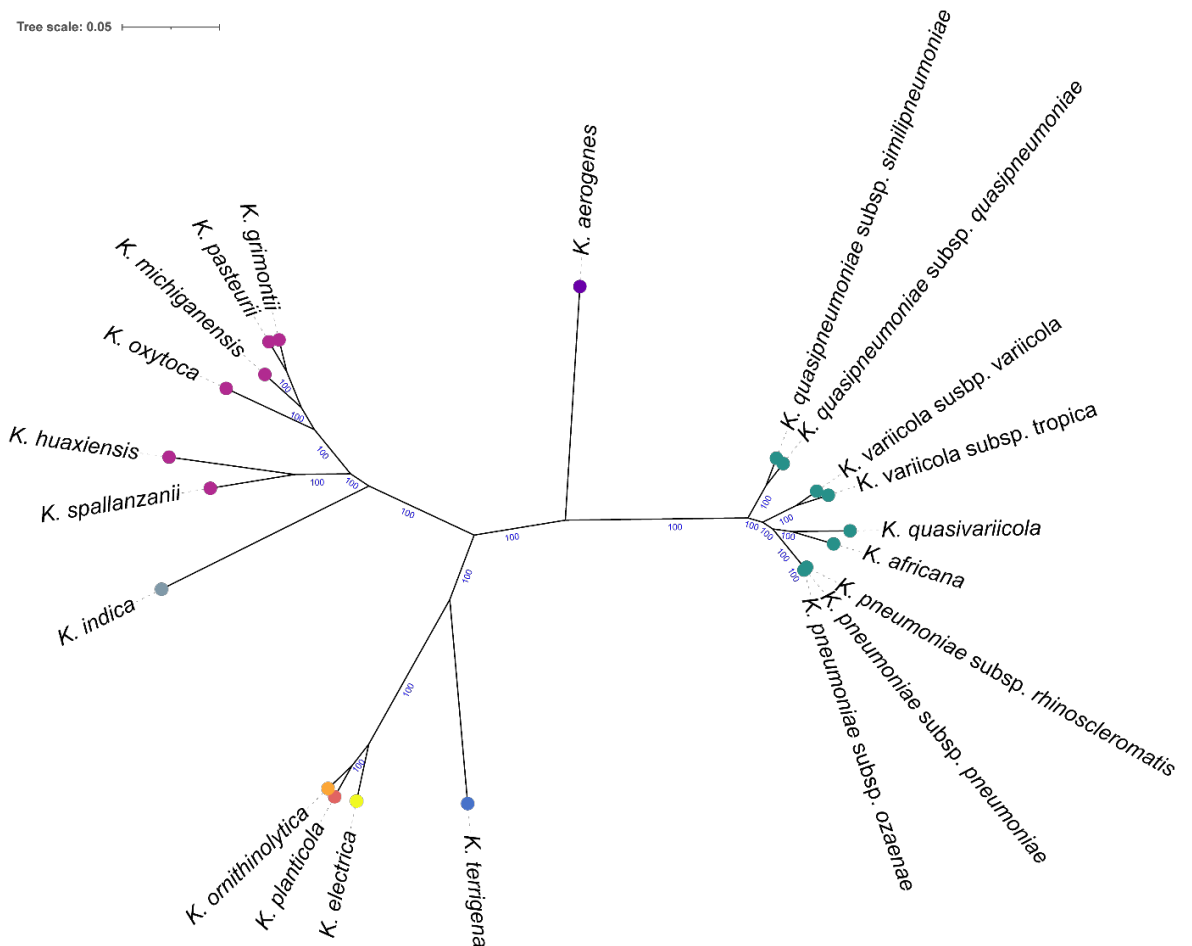
Figure 1.1. Population structure of *Klebsiella* spp. Phylogenetic relationship of members of the genus *Klebsiella*, including *K. pneumoniae* species complex (green) and *K. oxytoca* species complex (pink). Individual species are given as a single colour. Maximum-likelihood phylogeny of core genome alignment with ultrafast bootstrapping generated with IQ-TREE. Tree rendering was performed with iTOL. Bootstrap support values indicated on branches when ≥ 70. Tree is unrooted. Scale bar represents number of nucleotide substitutions per site.

Recently, *K. aerogenes* (formerly *Enterobacter aerogenes*) was re-assigned to the genus *Klebsiella* as genomic analysis revealed it was more closely related to *K. pneumoniae* than to *Enterobacter* (Malek *et al.*, 2019). *K. aerogenes* is phenotypically distinct from other *Klebsiella* spp. as it is the only motile species (Merla *et al.*, 2019).

Previously, *K. electrica*, K. *ornithinolytica*, *K. planticola*, and *K. terrigena* were transferred from *Klebsiella* to a new genus, *Raoultella*, rendering the *Klebsiella* as non-monophyletic; that is, after re-assignment the group *Klebsiella* was interspersed with members of the new genus *Raoultella* (McLennan, 2010; Wyres *et al.*, 2020; Ma *et al.*, 2021). However, recent phylogenetic data confirmed that these species should be re-assigned to the genus *Klebsiella* (Ma *et al.*, 2021). Another species of *Klebsiella* has also been identified, *K. indica*, which together with the former *Raoultella* spp. and *K. variicola*, have been described as "plant-associated," of which little is known about their pathogenicity in humans (Appel *et al.*, 2021).

Table 1.1. *Klebsiella pneumoniae* and *Klebsiella oxytoca* species complex and their phylogroups.

Species	Phylogroup	Reference
Klebsiella pneumoniae complex		
K. pneumoniae (*sensu stricto*)		(Brisse and Verhoef, 2001; Wyres *et al.*, 2020; Wyres *et al.*, 2020)
K. pneumoniae subsp. *pneumoniae*	Kp1	-
K. pneumoniae subsp. *rhinoscleromatis*	Kp1	-
K. pneumoniae subsp. *ozaenae*	Kp1	-
K. quasipneumoniae		(Brisse *et al.*, 2014; Wyres *et al.*, 2020)
K. quasipneumoniae subsp. *quasipneumoniae*	Kp2	-
K. quasipneumoniae subsp. *similipneumoniae*	Kp4	-
K. variicola		(Rosenblueth *et al.*, 2004; Maatallah *et al.*, 2014; Rodrigues *et al.*, 2019; Wyres *et al.*, 2020)
K. variicola subsp. *variicola*	Kp3	-
K. variicola subsp. *tropica*	Kp5	-
K. quasivariicola	Kp6	(Long *et al.*, 2017; Wyres *et al.*, 2020)
K. africana	Kp7	(Rodrigues *et al.*, 2019; Wyres *et al.*, 2020)
Klebsiella oxytoca complex		
K. oxytoca (*sensu stricto*)	Ko2	(Brisse and Verhoef, 2001)
K. grimontii	Ko6	(Passet and Brisse, 2018)
K. michiganensis	Ko1	(Saha *et al.*, 2013)
K. spallanzanii	Ko3	(Merla *et al.*, 2019)
K. pasteurii	Ko4	(Merla *et al.*, 2019)
K. huaxiensis	Ko8	(Hu *et al.*, 2019)

Preliminary identification is usually performed with ribosomal multilocus sequence typing (rMLST) that analyses the 53 *rps* subunit genes (Jolley *et al.*, 2012), and taxonomic assignment is often confirmed with average nucleotide identity (ANI) and phylogeny. Further classification of *Klebsiella* isolates is typically performed by MLST of seven housekeeping genes (*rpoB*, *gapA*, *mdh*, *pgi*, *phoE*, *infB*, *tonB*), and is used to describe their clonal relationships and population structure into distinct sequence types (STs) (Diancourt *et al.*, 2005; Herzog *et al.*, 2014). STs are grouped into clonal complexes (CCs) by their similarity to a predominant genotype, whereby each other member shares a minimum of five alleles with at least one other in the group (Feil and Spratt, 2001; Herzog *et al.*, 2014). For *K. pneumoniae*, the CC258 consists of several STs, where ST258 is the most globally dominant clone (Chen *et al.*, 2014; Pitout *et al.*, 2015). Whilst less well established, MLST

studies have also identified *K. oxytoca* CC2 as globally disseminated (Herzog *et al.*, 2014; Izdebski *et al.*, 2015).

1.1.2 Clinical significance of *Klebsiella* spp.

Human mucosal carriage rates of *Klebsiella* spp. can vary considerably, with gastrointestinal and nasopharyngeal rates at an estimated 10% and 5%, respectively (Podschun and Ullman, 1998; Högenauer *et al.*, 2006; Gorrie *et al.*, 2018; Wyres *et al.*, 2020). These rates increase significantly within healthcare settings however (Podschun and Ullman, 1998). Moreover, mucosal carriage is a major risk factor for developing disease, making *Klebsiella* spp. important opportunistic pathogens in humans, where they primarily affect the elderly, neonates, and the immunocompromised (Podschun and Ullman, 1998; Paczosa and Mecsas, 2016).

Klebsiella spp. are responsible for a range of infections, including urinary tract infections (UTIs), wound and surgical site infections, pneumonia, and bacteraemia, which can lead to sepsis (Podschun and Ullman, 1998; Holt *et al.*, 2015; Paczosa and Mecsas, 2016). *K. pneumoniae* and *K. oxytoca* account for approximately 75% and 20% of recovered isolates, respectively (Broberg *et al.*, 2014).

The ability of *K. pneumoniae* to rapidly disseminate within healthcare settings has earned it notoriety as a major cause of healthcare-associated infections (HAIs), accounting for approximately 30% of all Gram-negative infections in such settings (Holt *et al.*, 2015; Navon-Venezia *et al.*, 2017; Bengoechea and Sa Pessoa, 2019). Association between *K. pneumoniae* and medical devices is well documented: colonisation of indwelling catheters and ventilator circuits can cause catheter-associated UTIs, and ventilator-associated pneumonia, respectively (Podschun and Ullman, 1998; Broberg *et al.*, 2014; Moradigaravand *et al.*, 2017). Moreover, *K. pneumoniae* infections can progress to bacteraemia, where it is second only to *Escherichia coli* in UTI-associated bacteraemia within healthcare settings (UK Health Security Agency, 2019). Together, these infections account for a significant proportion of all HAIs in the UK, placing significant strain on resources (NICE, 2014; UK Health Security Agency, 2019).

Whilst most classical strains of *K. pneumoniae* are strongly associated with antibiotic resistance and healthcare settings, a phenotypically distinct form, hypervirulent *K. pneumoniae* (HvKP) has emerged in Southeast Asia over the last three decades (Paczosa and Mecsas, 2016; Wyres *et al.*, 2019). These HvKP strains additionally affect immunocompetent individuals, causing severe community-acquired infections including pyogenic liver abscesses and meningitis (Broberg *et al.*, 2014; Paczosa and Mecsas, 2016). In contrast to classical *K. pneumoniae* strains, HvKP strains are not usually associated with antibiotic resistance and their ability to cause a more clinically severe disease state is due to the presence of specific virulence factors (Paczosa and Mecsas, 2016; Wyres *et al.*,

2020). Whilst this demonstrates the divergence of the two types of *K. pneumoniae* into distinct subpopulations, there have been reports of HvKP strains carrying antibiotic resistance determinants, indicating global collaborative effort is needed to address this urgent threat to public health (Holt *et al.*, 2015; Roulston *et al.*, 2018; Wyres *et al.*, 2019).

K. oxytoca is also responsible for a similar range of infections as *K. pneumoniae*, but is most often associated with diseases affecting the colon, including antibiotic-associated haemorrhagic colitis (AAHC) (Högenauer *et al.*, 2006) and necrotising enterocolitis (NEC) (Chen *et al.*, 2020). AAHC is induced following antibiotic therapy, where overgrowth of antibiotic-resistant *K. oxytoca* results in bloody diarrhoea and abdominal pain (Högenauer *et al.*, 2006; Darby *et al.*, 2014; Hering *et al.*, 2019; Unterhauser *et al.*, 2019). AAHC is differentiated from antibiotic-associated colitis by the absence of toxigenic *Clostridioides difficile* and resolution of symptoms once antibiotic therapy is stopped (Högenauer *et al.*, 2006; Hering *et al.*, 2019; Unterhauser *et al.*, 2019). Approximately half of all *K. oxytoca* gut isolates are enterotoxigenic, a characteristic that contributes to their pathogenicity in AAHC (Högenauer *et al.*, 2006; Darby *et al.*, 2014; Unterhauser *et al.*, 2019). A recent study however, found that an enterotoxin-independent mechanism also drives pathogenicity of *K. oxytoca* in AAHC, and this may have clinical implications for enterotoxin-negative carriers (Hering *et al.*, 2019).

NEC is most commonly associated with preterm infants, particularly those of very low birthweight (Thompson and Bizzarro, 2008). Preterm infants often receive intensive antibiotic therapy which displaces the protective species of the gut microbiota and selects for antibiotic-resistant pathogens, including *Klebsiella* spp. (Gibson *et al.*, 2016; Chen *et al.*, 2020; Leggett *et al.*, 2020; Sequeira *et al.*, 2020). This enrichment of the infant gut with *Klebsiella* and other *Enterobacteriaceae* contributes to NEC development, which can progress to sepsis (Thompson and Bizzarro, 2008; Alcon-Giner *et al.*, 2020; Chen *et al.*, 2020). Indeed, Alcon-Giner *et al.* (2020) found that supplementation of the gut microbiota with *Bifidobacterium* and *Lactobacillus* led to a reduction of pathogens in preterm infants.

The clinical significance of other *Klebsiella* spp. has yet to be fully elucidated, but emerging evidence suggests that lesser-known species may account for a higher rate of infections, owing to improved detection and speciation methods. Recent studies have found that higher mortality rates in bacteraemic patients were associated with *K. variicola* isolates when compared to *K. pneumoniae,* and this association was independent of virulence factors and antibiotic resistance determinants (Seki *et al.*, 2013; Maatallah *et al.*, 2014). Moreover, the first clinical case of a *K. michiganensis* harbouring a *K. pneumoniae* carbapenemase (KPC) further emphasises how little is known regarding antibiotic resistance and virulence patterns among other *Klebsiella* spp. (Seiffert *et al.*, 2019). It remains to be determined as to whether these patterns influence bacterial inhabitation of a niche environment (Holt *et al.*, 2015;

Wyres *et al.*, 2019). Therefore, the importance of correct identification for epidemiological and outbreak studies cannot be underestimated.

1.1.3 Virulence factors of *Klebsiella* spp.

Klebsiella spp. possess a range of virulence factors that contribute to their pathogenicity (**Figure 1.2**), and are predominantly involved in colonisation and immune evasion (Li *et al.*, 2014; Paczosa and Mecsas, 2016). The capsule of *Klebsiella* spp. is the best described virulence factor and is responsible for the mucoid phenotype that typifies the genus (Lawlor *et al.*, 2005). The capsular polysaccharides (CPS) that make up the capsule have historically been used to type *Klebsiella*, resulting in a strain-specific K-antigen profile (Podschun and Ullman, 1998; Paczosa and Mecsas, 2016). There are more than 75 known capsular types, where K1 and K2 are most commonly associated with clinical *K. pneumoniae* infections (Broberg *et al.*, 2014; Pan *et al.*, 2015). In the UK, the most commonly isolated capsule types include KL2, KL3, KL15, KL20, KL24, KL45, KL51, KL64, KL81, KL102, and KL105 (Mentasti *et al.*, 2023; Centre for Genomic Pathogen Surveillance, 2024). CPS are integral to immune evasion, mainly by inhibiting phagocytosis and complement-mediated lysis, leading to an impaired inflammatory response during infection (Li *et al.*, 2014; Paczosa and Mecsas, 2016; Wyres *et al.*, 2016; Holt *et al.*, 2020). Indeed, acapsular *K. pneumoniae* strains are attenuated in murine models (Lawlor *et al.*, 2005; Fung *et al.*, 2011).
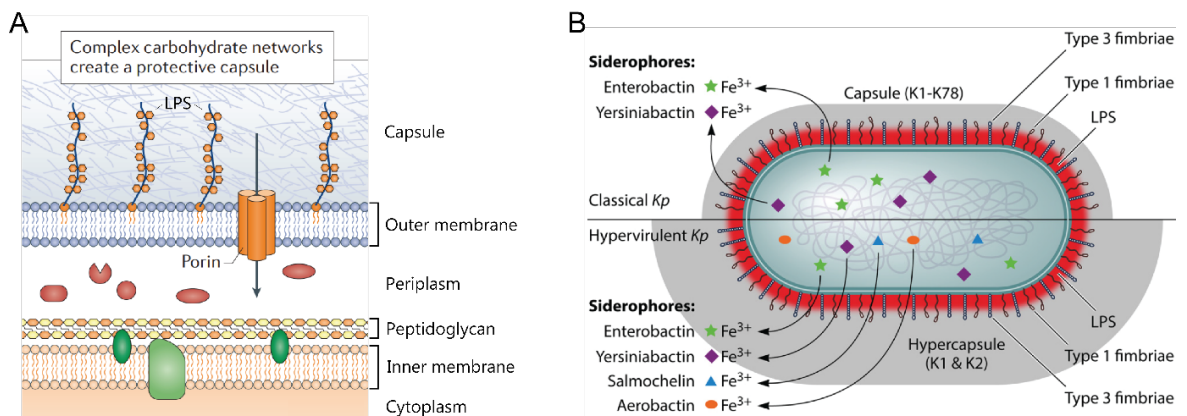
Figure 1.2. Structural composition of the cell wall of *Klebsiella* spp. (A) Schematic representation of outer membrane components that contribute to virulence and intrinsic antibiotic resistance among of *Klebsiella* spp. Adapted from Tommasi *et al.* (2015). **(B)** Schematic representation of the major virulence factors and their differences in classical and hypervirulent strains of *K. pneumoniae*. Reproduced from Paczosa and Mecsas (2016).

As is typical for Gram-negative bacteria, the outer membrane (OM) of *Klebsiella* spp. comprises a lipopolysaccharide (LPS) outer leaflet that is important in maintaining membrane integrity (Paczosa and Mecsas, 2016). LPS comprises a core oligosaccharide, an O-antigen, and a lipid A moiety, with the latter being responsible for toxicity during infection, known as endotoxin (Podschun and Ullman, 1998; Hsieh *et al.*, 2012). LPS

exhibits both pro- and anti-inflammatory activities during *K. pneumoniae* infection (Bengoechea and Sa Pessoa, 2019; Holt *et al.*, 2020). O-antigen facilitates colonisation and inhibits complement-mediated lysis, and whilst lipid A can inhibit antimicrobial peptide activity, it is strongly immunogenic (Hsieh *et al.*, 2012). Indeed, endotoxin is a contributing factor to overstimulation of the immune response and poor outcomes during infection, particularly in cases of sepsis (Lepper *et al.*, 2002; Dufour *et al.*, 2019).

To ensure their continued growth and survival, *Klebsiella* spp. produce a range of siderophores allowing them to sequester iron that is usually bound by human proteins, such as transferrin and lactoferrin (Tarkkanen *et al.*, 1992; Holt *et al.*, 2015; Holden *et al.*, 2016). Siderophore production among the *Klebsiella* spp. has been best studied in *K. pneumoniae*, with clear differences in types and expression levels between classical and hypervirulent strains (Roulston *et al.*, 2018).

Enterobactin is the predominant siderophore in all *K. pneumoniae* strains, with some classical strains also producing yersiniabactin, both of which are inhibited by human-derived lipocalin-2 and transferrin, respectively (Holden *et al.*, 2016). As well as upregulating the expression of enterobactin and yersiniabactin, HvKP strains have acquired additional siderophores salmochelin and aerobactin (Roulston *et al.*, 2018; Wyres *et al.*, 2019). These latter siderophores are rarely encountered in classical strains (Holt *et al.*, 2015; Paczosa and Mecsas, 2016), and they significantly increase the pathogenic potential of HvKP (Holden *et al.*, 2016).

Klebsiella spp. express phenotypically distinct type 1 and type 3 fimbriae, which are differentially expressed depending on the niche environment (Tarkkanen *et al.*, 1992; Stahlhut *et al.*, 2012). Both types of fimbriae are conserved among many *Enterobacteriaceae* and are implicated in mucosal colonisation and attachment to human cells during infection, biofilm formation, adhesion to medical devices, and immune activation (Tarkkanen *et al.*, 1992; Stahlhut *et al.*, 2012).

1.1.4 *Klebsiella* and antimicrobial resistance

Antimicrobial resistance (AMR) is a natural evolutionary phenomenon that enables bacteria and other microorganisms to grow in the presence of a particular compound to which it was designed to inhibit (Hiltunen *et al.*, 2017). Whilst there are no standardised definitions for multidrug, extensively-drug, and pan-drug resistance (MDR, XDR, and PDR, respectively) (Magiorakos *et al.*, 2012; Basak *et al.*, 2016), general definitions used throughout this project are:

- MDR: resistance to antibiotics representative of three or more classes.
- XDR: resistance to at least one antibiotic in all except two classes.
- PDR: resistance to all antibiotics.

An antibiotic class is defined as a group of antibiotics with similar chemical structure that exhibit highly similar mechanism of action (Malmir *et al.*, 2018; Hutchings *et al.*, 2019). Increasing antibiotic resistance among *Klebsiella* isolates is largely driven by decades of antibiotic misuse in medicine and agriculture, whereby exposure to sub-inhibitory concentrations of antibiotics has driven the selection for antibiotic resistance determinants (Hiltunen *et al.*, 2017; WHO, 2017). The emergence of MDR among *Klebsiella* spp. has severely reduced therapeutic options, resulting in higher morbidity and mortality (Holt *et al.*, 2015; Theuretzbacher, 2017).

Klebsiella spp. are increasingly recognised as prolific accumulators and disseminators of antibiotic resistance determinants, and these can either be chromosomally-encoded or plasmid-derived (Holt *et al.*, 2015; Gorrie *et al.*, 2018). The human gut microbiota may also present a concentrated reservoir of antibiotic resistance determinants that are readily shuttled among pathogenic members (Holt *et al.*, 2015; Chen *et al.*, 2020; Sequeira *et al.*, 2020). These factors also contribute to the dominance of *Klebsiella* spp. within healthcare settings.

Klebsiella spp. are intrinsically resistant to some antibiotics due their cell wall structure, prominent capsule (**Figure 1.2**) and ability to produce biofilms (Tommasi *et al.*, 2015; Paczosa and Mecsas, 2016). Acquired antibiotic resistance determinants in *Klebsiella* spp. include an array of β-lactamase enzymes, including extended-spectrum β-lactamases (ESBLs) that hydrolyse multiple β-lactam antibiotics (Queenan and Bush, 2007), including carbapenems, which are considered a last line of defence against MDR pathogens (**Table 1.2**) (Pitout *et al.*, 2015; Paczosa and Mecsas, 2016; Theuretzbacher, 2017). Indeed, carbapenem-resistant *Enterobacteriaceae* are associated with recalcitrant infections, as is highlighted by the World Health Organization's priority pathogens list (WHO, 2017). Moreover, the prevalence of carbapenem-resistant *K. pneumoniae* (CRKP) is increasing at an alarming rate and these strains are now endemic in Italy and Greece (Theuretzbacher, 2017; Cassini *et al.*, 2019). Whilst rare in the UK, carbapenem-resistant infections are increasing, with endemic rates observed in Manchester (Donker *et al.*, 2017; Decraene *et al.*, 2018). Carbapenems are a last resort antibiotic for recalcitrant *Klebsiella* infections, but the last resort antibiotic for Gram-negative bacteria, colistin has been increasingly prescribed due to the emergence of MDR and XDR clones (Sabnis *et al.*, 2021). The drug is not without consequences however, where physicians need to balance the risk of infection and sepsis, or nephrotoxicity that can lead to permanent kidney damage (Sabnis *et al.*, 2021).

Table 1.2. Summary of β-lactamases among Enterobacteriaceae with substrate and clavulanic acid inhibition profiles. Adapted from (Queenan and Bush, 2007).

Molecular class	Enzyme	Penicillins	Early cephalosporins	Extended-spectrum cephalosporins	Aztreonam	Carbapenems	Clavulanic acid* inhibition	Source
Penicillinases								
A	OXY	+	+	±	±	-	variable	(Hæggman *et al.*, 2004; Fevre *et al.*, 2005; Fevre *et al.*, 2005; González-López *et al.*, 2009)
	OKP	+	±	-	-	-	y	
ESBLs[a]								
A	SHV	+	+	±	±	-	y	(Hæggman *et al.*, 2004; Cantón *et al.*, 2012; Tooke *et al.*, 2019; Sawa *et al.*, 2020)
	TEM	+	+	±	±	-	y	
	CTX-M	+	+	+	+	-	y	
AmpC enzymes								
C	CMY	+	+	±	±	-	variable	(Jacoby, 2009; Tooke *et al.*, 2019; Sawa *et al.*, 2020)
	AmpC	+	+	±	±	-	variable	
Carbapenemases								
A	IMI	+	+	+	+	+	y	(Queenan and Bush, 2007; Tooke *et al.*, 2019; Sawa *et al.*, 2020)
	SME	+	+	±	+	+	y	
	KPC	+	+	+	+	+	y	
	GES	+	+	+	-	±	y	
B1	IMP	+	+	+	-	+	n	(Queenan and Bush, 2007; Tooke *et al.*, 2019; Sawa *et al.*, 2020)
	VIM	+	+	+	-	+	n	
	GIM	+	+	+	-	+	n	
	NDM	+	+	+	-	+	n	
D	OXA	+	+	±	-	±	variable	(Queenan and Bush, 2007; Maurya *et al.*, 2017; Tooke *et al.*, 2019; Sawa *et al.*, 2020)

*, clavulanic acid and tazobactam are both β-lactamase inhibitors with similar spectrum of activity; [a], all ESBLs have early cephalosporin activity and some are not considered "true" ESBLs, including SHV-1, SHV-11, and TEM-1. All other SHV and TEM variants are considered ESBLs, therefore aztreonam activity is for "true" ESBL variants only.

Carbapenemase enzymes confer resistance to a range of β-lactams depending on their molecular class (**Table 1.2**). For example, class A enzymes such as KPC can hydrolyse extended-spectrum β-lactams, carbapenems, and aztreonam (Queenan and Bush, 2007; Pitout *et al.*, 2015). Conversely, class D enzymes such as OXA carbapenemases demonstrate variable activity against extended-spectrum β-lactams and carbapenems, with no activity against aztreonam (Queenan and Bush, 2007; Pitout *et al.*, 2015). Moreover, carbapenemases are distinct in their inhibition profiles: class A enzymes are inhibited by β-lactamase inhibitors such as clavulanic acid and tazobactam, whereas class B metallo-β-lactamases are not (Queenan and Bush, 2007).

Outer membrane porin defects have also been shown to contribute to the MDR phenotype among CRKP strains. Porins are trimeric proteins that span the OM of Gram-negative bacteria, and are important in the maintenance of cellular permeability and structural integrity (Tsai *et al.*, 2011; Wong *et al.*, 2019). Additionally, porins contribute to the diffusion of antibiotics across the OM and into the periplasmic space (Pagès *et al.*, 2008; James *et al.*, 2009; Wong *et al.*, 2022). *Klebsiella* spp. express a number of different porins, including OmpK35 and OmpK36, which are homologs of OmpF and OmpC in *E. coli*, respectively (James *et al.*, 2009; Tsai *et al.*, 2011). Selection for porin mutations during antibiotic therapy most commonly manifests with reduced functionality or complete loss of OmpK35 and/or OmpK36 (Pagès *et al.*, 2008; Tsai *et al.*, 2011; Wong *et al.*, 2019; Wong *et al.*, 2022).

The globally dominant *K. pneumoniae* ST258 is a CRKP strain capable of acquiring an array of β-lactamase genes (Chen *et al.*, 2014; Bowers *et al.*, 2015), with porin modifications additionally contributing to the MDR phenotype (Pagès *et al.*, 2008; Tsai *et al.*, 2011; Bowers *et al.*, 2015; Buckner *et al.*, 2018; Wong *et al.*, 2019). Genomic analysis of ST258 isolates revealed truncation of OmpK35 caused by a frameshift mutation that results in a premature stop codon (Bowers *et al.*, 2015; Wong *et al.*, 2019). This truncation produces a non-functional protein that is highly conserved among these isolates (Bowers *et al.*, 2015). Whilst OmpK36 mutations are more variable, a key mutation includes a glycine-aspartate insertion within the extracellular loop 3, creating pore constriction that prevents antibiotic influx (Bowers *et al.*, 2015; Wong *et al.*, 2019). Moreover, synonymous mutations have been described that disrupt translation of *ompK36* and are selected for in response to antibiotic therapy (Wong et al., 2022). Reduced expression or complete loss of OmpK35 and OmpK36 porins have shown reduced virulence *in vivo*, suggesting increased antibiotic resistance results in a fitness cost. Indeed, in a murine model of severe *K. pneumoniae* lung infection, OmpK35 and OmpK36 double deletion mutants were attenuated when compared to wild-type infection (Tsai *et al.*, 2011; Wong *et al.*, 2019). Moreover, some attenuated mutations have limited capacity for global expansion (Wong *et al.*, 2022). Despite the associated fitness cost of porin modifications, the global success of ST258 still remains elusive (Pagès *et al.*, 2008; Bowers *et al.*, 2015; Holt *et al.*, 2015; Bengoechea and Sa Pessoa, 2019; Wong *et al.*, 2019).

1.2 Bacteriophage diversity and biology

Bacteriophages (phages) are viruses that infect bacteria and are the most abundant and diverse biological entities on Earth (Hendrix, 2002; Mushegian, 2020). They have been found in almost all microbial niches, where they are integral to bacterial ecology, metabolism, and diversity by driving evolution through natural selection and shuttling of genetic material (Mavrich and Hatfull, 2017; Ofir and Sorek, 2018; Chiang *et al.*, 2019; Al-Shayeb *et al.*, 2020; Dion *et al.*, 2020; Mushegian, 2020).

Most of our understanding of bacteriophage biology has originated from studies performed on a few model phages that infect *E. coli*, collectively known as the "T" phages and phage Lambda. These phages belong to the class *Caudoviricetes*, otherwise known as tailed (dsDNA) phages (**see section 1.2.1**). Relatives of the "T" phages in particular are overrepresented in many phage isolation and therapy studies. Phages contribute significantly to the pool of novel genetic material accessible to their bacterial hosts, thereby driving their genetic diversity and evolution (Khan *et al.*, 2021). Despite their importance, compared to bacteria, little is known about how phages (outside of the model "T" and Lambda phages) interact with their bacterial hosts, specifically with regards to the phenotype-genotype link (Clokie *et al.*, 2011; Maffei *et al.*, 2021).

1.2.1 Phage diversity and classification

Phage classification has historically been based on nucleic acid composition and morphology (Ackermann, 2005). Nucleic acid composition of phages can be DNA or RNA, either double-stranded (ds) or single-stranded (ss) (Ofir and Sorek, 2018; Romero-Calle *et al.*, 2019; Koonin *et al.*, 2020), while morphologically, phages can be distinguished as tailed (dsDNA), filamentous or rod-like (ssDNA), icosahedral/non-tailed (ssRNA, ssDNA, or dsDNA), and pleiomorphic (dsDNA) (**Figure 1.3**) (Ackermann, 2005; Hay and Lithgow, 2019; Barylski *et al.*, 2020; Dion *et al.*, 2020; Callanan *et al.*, 2021). Whilst tailed phages, i.e. those in the class *Caudoviricetes*, represent the majority of those isolated, methodologies may favour their isolation due to their stability, suggesting that RNA phages are largely underrepresented (Callanan *et al.*, 2018; Grasis, 2018). Indeed, as most studies have focused on the tailed phages, there is a paucity of published literature on the biology of other phages (Callanan *et al.*, 2018). However, advancements in genomic and metagenomic analyses have led to an expansion in the number of ssRNA viruses being discovered, and a subsequent overhaul of their taxonomic classification (Callanan *et al.*, 2021).

Bacteriophage, and more generally virus, taxonomy is complicated by genomic mosaicism of certain groups of phages and the lack of a universal marker gene (Ofir and Sorek, 2018; Dion *et al.*, 2020). Previously, phages and other viruses were classified based on their nucleic acid structure, known as the Baltimore Class (ICTV Executive Committee, 2020;

Koonin *et al.*, 2020). Recently, however, viruses have been classified into realms, which represent their independent evolutionary origins (Gorbalenya *et al.*, 2019; ICTV Executive Committee, 2020; Koonin *et al.*, 2020). This project will focus solely on members within the realm *Duplodnaviria*, class *Caudoviricetes*, also known as the tailed (dsDNA) phages.
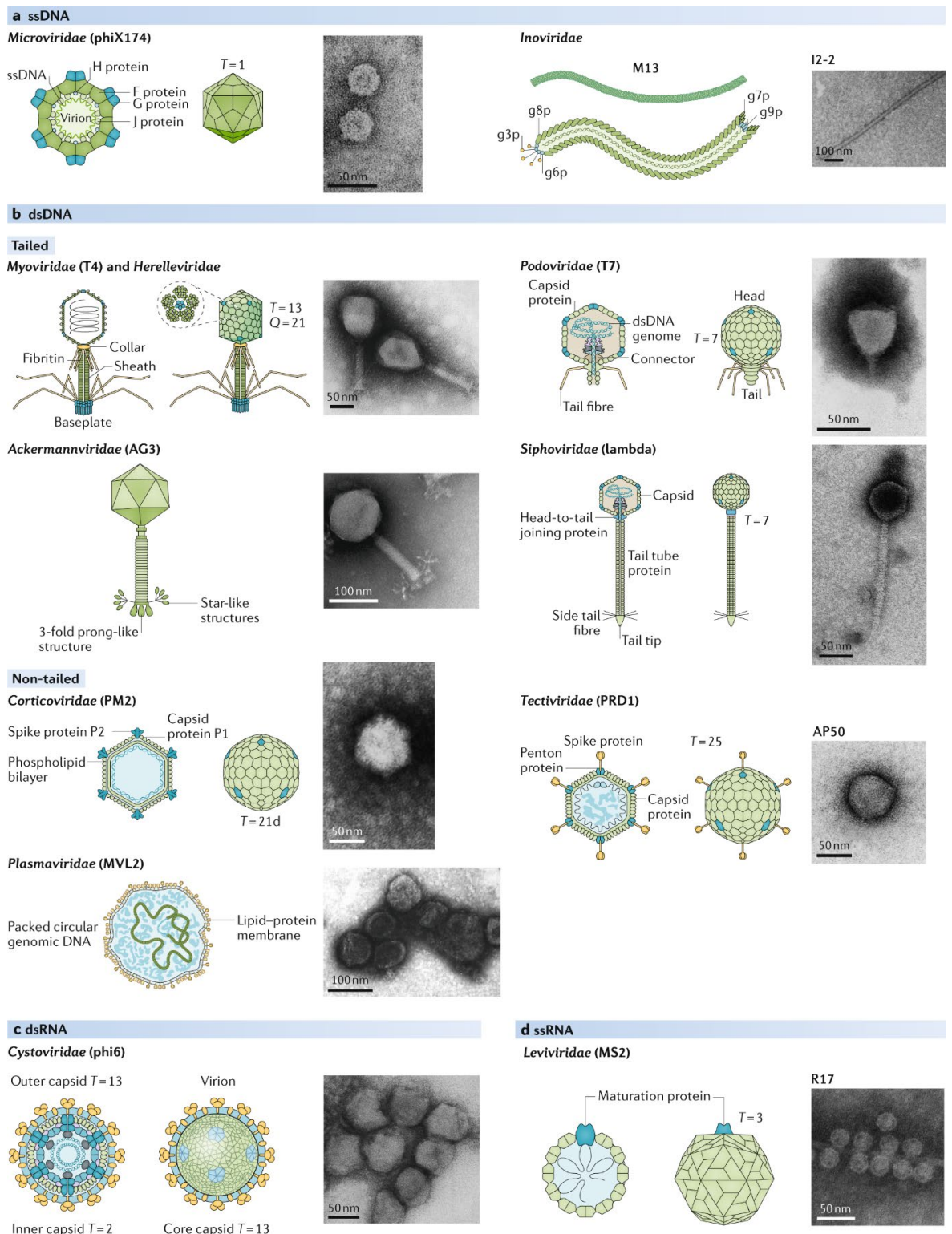
Figure 1.3. Schematic representation of morphological descriptions and genomic composition of phages as of 2020. An example TEM image of an exemplar phage is given for each family. **(a)** Morphology of the families *Microviridae* and *Inoviridae* as examples of ssDNA phages. **(b)** Morphology of the dsDNA phages including tailed (*Caudoviricetes*) and non-tailed phages. **(c)** Morphology of dsRNA phages, with the family *Cystoviridae* as the only recognised family in the class *Vidaverviricetes*. **(d)** Morphology of ssRNA phages with

13

family *Fiersviridae* (formerly *Leviviridae*) as an exemplar of six families in the class *Leviviricetes.* Reproduced from Dion *et al.* (2020).

Previously, tailed phages were assigned to a morphology-based family-rank taxonomic structure based on the tail structure: *Myoviridae* (long, contractile tails), *Podoviridae* (short non-contractile tails), and *Siphoviridae* (long, non-contractile tails). However, certain groups of morphologically distinct phages shared genomic homology, particularly between temperate phages related to *Escherichia* phage Lambda (siphovirus) and *Salmonella* phage P22 (podovirus), whereas other phages belonging to the same morphology-based family shared no genomic similarity at all (Rohwer and Edwards, 2002; Ackermann, 2005; Turner *et al.*, 2021). This has led to the dismantling of the morphology-based families (Turner *et al.*, 2023). Currently, there are 41 families, over 1,300 genera, and over 4,000 species that have been extracted from the now obsolete morphology-based families *Myoviridae*, *Podoviridae*, and *Siphoviridae* (Adriaenssens *et al.*, 2018; Adriaenssens *et al.*, 2020; Barylski *et al.*, 2020; Barylski *et al.*, 2020).

1.2.2 Phage infection cycle

A discussion of the infection cycle of all phage types is beyond the scope of this project, and this section will primarily focus on the infection cycle of model phages of the class *Caudoviricetes*.

Like all viruses, phages are obligate parasites capable of hijacking the host replicative machinery to produce phage progeny. The first stage of infection requires adsorption through interaction with host receptors. This is determined by receptor binding proteins (RBPs) found on various tail structures, including long tail fibres, short tail fibres, and/or tail spikes (Nobrega *et al.*, 2018; Ofir and Sorek, 2018; Gonzalez-Serrano *et al.*, 2023). The nature of RBP attachment to bacterial cell surface components results in a cycle of bacterial modification of surface components and drives evolution of intracellular anti-phage defence mechanisms, which in turn, drives co-evolution of phage-mediated counter-defence mechanisms (Chevallereau *et al.*, 2022; Gonzalez-Serrano *et al.*, 2023). This co-evolution predisposes phages to rapidly modify their RBP through mutation and/or recombination (Kutter *et al.*, 2005; Chevallereau *et al.*, 2022; Gonzalez-Serrano *et al.*, 2023).

Targeted receptors among Gram-negative bacteria may include CPS, LPS, porins or other OM proteins, pili, or flagella (Kutter *et al.*, 2005; Bertozzi Silva *et al.*, 2016; Gonzalez-Serrano *et al.*, 2023). Additionally, degradation of CPS and/or biofilms by phage enzymes such as capsular depolymerases is often necessary for access and adsorption to bacterial surface components (Cai *et al.*, 2019). After encountering a host, adsorption proceeds by both reversible and irreversible attachment of specific surface components (Guttman *et al.*, 2005; Bertozzi Silva *et al.*, 2016; Maffei *et al.*, 2021). Reversible attachment involves weak associations between RBPs and the host, often sugar moieties on cell surface components

such as glycans of the O-antigen and enterobacterial common antigen (ECA) (Guttman *et al.*, 2005; Bertozzi Silva *et al.*, 2016; Maffei *et al.*, 2021). Irreversible attachment to specific receptors is often by different RBPs, triggering a conformational change that leads to ejection of phage DNA from the capsid into the host (Kutter *et al.*, 2005; Bertozzi Silva *et al.*, 2016; Nobrega *et al.*, 2018). Studies in the model *E. coli* phages T4 and Lambda found that reversible attachment appears to initiate 'receptor walking' along the bacterial cell, where they exhibit a preference for irreversible attachment at the bacterial poles (Molineux, 2006; Edgar *et al.*, 2008; Chatterjee and Rothenberg, 2012; Hu *et al.*, 2015). Moreover, reversible attachment is a prerequisite to infection, that increases the likelihood of encountering receptors for irreversible attachment (Baptista *et al.*, 2008; Chatterjee and Rothenberg, 2012; Bertozzi Silva *et al.*, 2016).

Specific morphotypes appear to target specific secondary receptors. For example, podoviruses almost exclusively target LPS components; whereas siphoviruses almost exclusively target porins (Molineux, 2006; Maffei *et al.*, 2021). The targeted receptors of the myoviruses however, appear to be dependent on the family or genus (Maffei *et al.*, 2021). For *Klebsiella* spp., it is unclear whether the capsule is utilised as a primary receptor for reversible attachment that results in its degradation to access receptors for irreversible attachment, or whether the capsule is a barrier to overcome to access both primary and secondary receptors (Molineux, 2006; González-García *et al.*, 2015; Latka *et al.*, 2017; Maffei *et al.*, 2021).

Following adsorption, phage DNA is injected into the host cytoplasm, where it may circularise to prevent host-induced enzymatic digestion (Guttman *et al.*, 2005). Phages can then adopt various life cycles including lytic, lysogenic, and pseudolysogenic (**Figure 1.4**). Which life cycle the phage proceeds down is dependent upon the type of phage, the expression of specific genes, and the conditions within the host cell (Guttman *et al.*, 2005).

The lytic life cycle proceeds after phage infection with expression of the phage genes necessary for host conversion into a phage factory, phage genome replication, and rapid assembly of virions, followed by lysis of the bacterial cell by specialised enzymes, releasing the progeny into the surrounding environment (Clokie *et al.*, 2011; Davies *et al.*, 2016). Early gene products reprogram the bacterial replicative machinery away from host-related functions, and late gene products direct phage particle production (Guttman *et al.*, 2005). All tailed phages possess an icosahedral capsid, into which DNA is packaged following capsid assembly (Guttman *et al.*, 2005). Once fully assembled, phage lysis can be initiated through the expression of lytic enzymes. Most tailed phages utilise a two-component system involving holins and endolysins, which disrupt membrane integrity and facilitate virion release (Guttman *et al.*, 2005; Žiedaitė *et al.*, 2005). Phages that replicate exclusively through the lytic cycle are known as virulent phages.
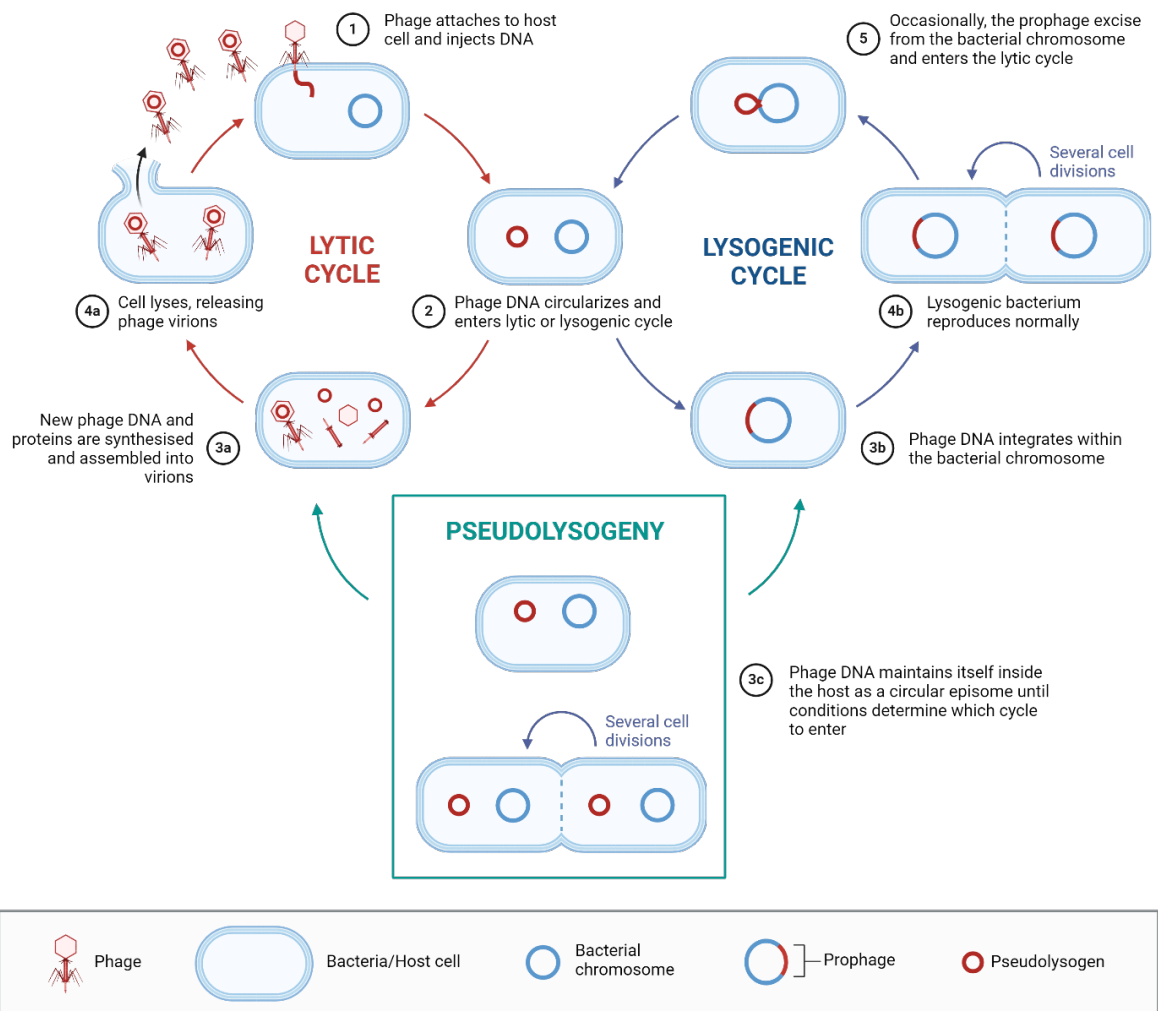
Figure 1.4. Schematic representation of the phage life cycles. Virulent phages enter the lytic cycle, proceed to assembly of phage progeny and rapid bacterial lysis. Temperate phages can enter either the lytic, lysogenic, or pseudolysogenic cycles, where the decision of which cycle to enter is dependent upon the conditions. Figure generated in BioRender.

Lysogeny is where temperate phages integrate their genetic material into the host genome without immediate lysis, which requires the expression of phage-encoded enzymes such as integrases (Nilsson and Ljungquist, 2006; Clokie *et al.*, 2011). The integrated phage genome is called a prophage and is replicated along with the host DNA. Prophages may switch to the lytic cycle when conditions are appropriate (Davies *et al.*, 2016; Ofir and Sorek, 2018). For example, host induction of the SOS response following DNA damage promotes the switch from the lysogenic to the lytic cycle (Little, 2006). A third, but less common lifestyle uses a pseudolysogenic life cycle, in which the phage genome maintains itself inside the host as a circularised episome (Abedon, 2008; Clokie *et al.*, 2011; Cenens *et al.*, 2013). Whilst the factors that determine pseudolysogeny are yet to be fully elucidated, it is thought to promote phage survival within the host until conditions allow entry to the lytic or lysogenic life cycle (Cenens *et al.*, 2013).

Whilst not as extensively studied as a group, there are a few model phages that have improved the understanding of non-tailed phages. For example, phiX174, a ssDNA *E. coli*

phage in the *Microviridae* family, also appears to demonstrate a lytic life cycle, but produces a single protein for escape (Fane *et al.*, 2006). Model ssRNA phages include MS2 and Qbeta, both of which infect *E. coli* (Callanan *et al.*, 2018). Indeed, almost all RNA phages identified to date infect only Gram-negative bacteria, where infection is often mediated by phage binding of host appendages, such as pili (van Duin and Tsareva, 2006; Callanan *et al.*, 2018).

1.3 Bacteriophage genomics

Genomics is the study of the entire genome of an organism or multiple organisms in reference to their structure, function, and evolution (Owen *et al.*, 2021). Understanding the complete genetic composition of organisms such as bacteria and their phages provides the necessary information on nucleotide and proteomic structure that underpins our understanding of infection biology, taxonomic classification, and treatment options in a phage therapy setting (Owen *et al.*, 2021; Turner *et al.*, 2021). Accurate genomics is integral to the use of phages as therapeutics. First, it is essential to identify potential virulence factors, toxins, and AMR genes that may be harboured on the phage genome, which could then be transferred to the bacterial host upon infection (Alavidze *et al.*, 2016; Abedon, 2017; Owen *et al.*, 2021; Turner *et al.*, 2021). Second, it is important to determine the phage life cycle: temperate phages are considered unsuitable for phage therapy and are identified by genes such as integrases (Abedon, 2017; Owen *et al.*, 2021; Turner *et al.*, 2021). Third, the uniqueness of each phage needed to be determined so that similar phages are not used in combination, as they may compete for the same host receptors, which would result in an antagonistic effect (Molina *et al.*, 2022). Lastly, understanding the genetic composition of the phage after each therapeutic application over time allows the evolutionary history to be tracked, enabling predictions to be made on the type of bacterial resistance that may develop, and the phage counter-resistance mechanisms that may emerge in response (Alavidze *et al.*, 2016). Moreover, understanding these types of interactions can better inform clinical decision-making, particularly with regard to using phages in combination with one another (i.e. in a "cocktail") and phage-antibiotic combinations (Alavidze *et al.*, 2016).

1.3.1 Bacteriophage sequencing and genome assembly

To determine the genetic composition of an organism, the genome needs to be sequenced. Underpinning genomics has been the advancement of high-throughput next generation sequencing (NGS) and third generation sequencing technologies has allowed increased power and resolution for investigating microbial genomes (Bansal and Boucher, 2019; Owen *et al.*, 2021). With improved NGS technologies has come improved bioinformatic tools for genome assembly and therefore improved genome quality, enabling a greater understanding of microbial biology in general.

Current sequencing technologies can broadly be separated into two main types: short-read and long-read sequencing. Short-read sequencing is more commonly used than long-read sequencing, and is often performed with Illumina technology, which generates short-reads of 50 to 300 base pairs (bp) (Adewale, 2020; Bharti and Grimm, 2021; Owen *et al.*, 2021; Cook *et al.*, 2024). Typically, short-read NGS involves ligation of adapters to fragmented DNA molecules, which are amplified and fluorescently labelled to distinguish each base from one another in the sequence (Adewale, 2020; Giani *et al.*, 2020; Owen *et al.*, 2021). Long-read sequencing can be performed by either Oxford Nanopore Technologies (ONT) or PacBio, where the former tends to be more cost effective, portable, and accessible (Adewale, 2020; Owen *et al.*, 2021). ONT involves sequencing a single DNA molecule through a 1.5 nm pore, negating the need for amplification-based methods (Adewale, 2020; Bharti and Grimm, 2021; Owen *et al.*, 2021). An electrical current is continually passed through the pore, which is disrupted as the DNA molecule is fed through so that each nucleotide provides a unique electrical signal (Adewale, 2020; Bharti and Grimm, 2021; Owen *et al.*, 2021). The development of long-read sequencing has allowed sequencing of much longer DNA fragments of approximately a few thousand bp to over 100,000 bp, where a whole phage genome can be contained within a single read (Bharti and Grimm, 2021; Owen *et al.*, 2021). Each type of sequencing technology has its advantages and disadvantages, and the use of one over another largely depends on the purpose. Whilst short-read sequencing is more cost effective for example, it cannot resolve repeat and hypervariable regions; whereas long-read sequencing is simpler to perform and more portable, but tends to be more error prone than short-read sequencing (Adewale, 2020; Amarasinghe *et al.*, 2020; Giani *et al.*, 2020; Bharti and Grimm, 2021; Lee *et al.*, 2021; Owen *et al.*, 2021; Cook *et al.*, 2024).

Bioinformatic tools generate a consensus sequence, known as an assembly, by "stitching" reads together into contigs, but can also be error-prone (Bharti and Grimm, 2021; Owen *et al.*, 2021). Moreover, sufficient read coverage is necessary to generate a reliable assembly, and is generally accepted as 100X (Owen *et al.*, 2021). Sequence read coverage is defined as the number of reads that align at a particular nucleotide location within the genome (Owen *et al.*, 2021).

An assembly can be contained within a single contig, but for much larger genomes such as bacterial genomes, the assembly may be spread over multiple or hundreds of contigs (Wick *et al.*, 2023; Cook *et al.*, 2024). Whether the assembly is in multiple contigs or not, without any further curation, the genomes are typically referred to as draft genomes because the nature of the sequencing and/or assembly can introduce errors that would not otherwise be present biologically (Bharti and Grimm, 2021; Owen *et al.*, 2021). Correction of sequencing and/or assembly errors is necessary to producing complete, accurate, and high-quality phage genomes (Owen *et al.*, 2021; Wick *et al.*, 2023). Error correction can be performed by mapping the reads back to the assembly and manually curating the assembly by eye, or

18

through use of polishing tools (Lee *et al.*, 2021; Owen *et al.*, 2021; Wick and Holt, 2022; Wick *et al.,* 2023). Polishing tools have their own algorithms that search for local misassemblies and inconsistencies in the assembly and corrects them using either short- or long-reads (Lee *et al.*, 2021; Owen *et al.*, 2021; Wick and Holt, 2022; Wick *et al.*, 2023). In an effort to improve the accuracy of long-read sequencing, ONT recently developed the Nanopore R10 flow cells to be used in conjunction with the R10 chemistry (Sereika *et al.*, 2022). The new chemistry has promised to minimise the need for assembly polishing by increasing read accuracy (Sereika *et al.*, 2022).

Assembly can be sequencing-type specific or can be a hybrid of the two types of sequencing (Wick *et al.*, 2017; Cook *et al.*, 2024). Many phage genomes for example, are assembled with the short-read sequencing data alone (Owen *et al.*, 2021; Cook *et al.*, 2024). In other circumstances, long-read sequencing and assembly is performed alone. A traditional hybrid assembly is where both long- and short-read sequencing data (i.e. the reads) are used in the assembly together (Cook *et al.*, 2024).

Phage genome assembly relies on specific algorithms of aligning reads to generate one or more contigs containing the genome. The subsequent assembly is randomly oriented and linearised. This means that for phages whose genes are unidirectional, i.e. are typically represented as left-to-right on the forward strand, an assembler may incorrectly orientate them on the reverse strand. Moreover, phages with defined ends such as those with direct terminal repeats (DTRs) (**see section 1.3.2**) may be linearised incorrectly, with the DTRs in the middle of the assembly and not flanking the genome. Incorrect orientation is a feature of both long- and short-read assemblers, whereas incorrect linearisation tends to be more problematic with short-read-only and traditional hybrid assemblies. Thus, it is important to identify the defined ends of phage genomes following assembly and reorientating them before database submission.

1.3.2 An overview of bacteriophage genome structure and packaging methods

Discussed below are the packaging methods used by a select few caudovirus phages that are relevant to the work in this project (**Figure 1.5**). A discussion of all packaging methods is beyond the scope of this project.

As phages assemble virion particles following infection, they need to package their DNA into the procapsid, which is facilitated by the small and large terminase subunits (TerS and TerL, respectively) (**Figure 1.5A**) (Casjens and Gilcrease, 2009; Li *et al.*, 2014; Merrill *et al.*, 2016; Khan *et al.*, 2021). The large terminase subunit (*terL*) is a functionally-conserved gene present in all tailed phages. However, there is considerable divergence in the TerS and TerL protein sequences between, and within, different phage families (Merrill *et al.*, 2016). This may reflect the remarkably diverse array of packaging methods that phages

employ to generate distinct termini (Black, 1989; Li *et al.*, 2014; Merrill *et al.*, 2016). Packaging method and the type of termini generated can thus be predicted in an unknown phage if there is shared protein sequence similarity in the TerL with another well-characterised phage (Merrill *et al.*, 2016).

Broadly, there are five types of packaging methods that generate distinct termini: DTRs, which can be short (~100-500 bp) or long (1000s bp); cohesive ends, which can have 5' or 3' overhangs at the termini; headful packaging that generates circularly permuted genomes; host ends; and covalent bound terminal proteins (Casjens and Gilcrease, 2009; Li *et al.*, 2014; Merrill *et al.*, 2016; Chiang *et al.*, 2019). In all packaging methods used, phages generate concatemers, often by circularisation of the genome following entry into the host, and rolling circle replication (Black, 1989; Casjens and Gilcrease, 2009; Li *et al.*, 2014; Merrill *et al.*, 2016; Chiang *et al.*, 2019). Concatemers are thus contiguous copies of the genome in a long, single DNA sequence. Terminases cleave the DNA after facilitating packaging, which may be at specific or non-specific sites, generating distinct termini (**Figure 1.5B-C**) (Casjens and Gilcrease, 2009; Li *et al.*, 2014; Merrill *et al.*, 2016; Chiang *et al.*, 2019).

Phage DNA packaging can also contribute to bacterial diversity through horizontal gene transfer (HGT), where phages package host DNA into capsids, transferring them to the next host in a process known as transduction (Chiang *et al.*, 2019). Such DNA can encode AMR determinants, metabolic genes, and/or virulence factors that confer a fitness advantage to the host (Matilla *et al.*, 2014; Chiang *et al.*, 2019; Borodovich *et al.*, 2022). For example, phage Lambda can transfer metabolic genes between its *E. coli* hosts through erroneous excision of genes flanking its *att* site (Chiang *et al.*, 2019; Borodovich *et al.*, 2022). Similarly, large quantities of DNA is packaged with the *Escherichia* phage Mu genome as cleavage occurs several 100 or 1000 bp up- or downstream of the termini (Kutter *et al.*, 2005). Moreover, phages can encode genes themselves (i.e. not transferred by transduction) that contribute to pathogenicity. A well-characterised example is the Shiga toxin, which is transferred to *E. coli* hosts by temperate phages (Yano *et al.*, 2023).

Genomes with short direct terminal repeats

The genome termini of phages belonging to the family *Autographiviridae* consist of short DTRs of up to 500 bp that flank the genome, and have been best studied in *Escherichia* phage T7 (Molineux, 2006). After rolling circle replication, concatemers are generated containing a single copy of the DTR at one end of the genome (Black, 1989; Molineux, 2006; Casjens and Gilcrease, 2009; Li *et al.*, 2014; Merrill *et al.*, 2016). The DNA is cleaved by the terminases and a second DTR is synthesised at the other end as the DNA is packaged into the procapsid with the assistance of portal proteins (**Figure 1.5B**) (Black, 1989; Casjens and Gilcrease, 2009; Li *et al.*, 2014; Merrill *et al.*, 2016). The role of the DTRs is yet to be fully elucidated, but they are thought to aid DNA replication and protect against

host-mediated degradation (Guttman *et al.*, 2005; Li *et al.*, 2014). DTRs can be identified through read mapping following WGS: they appear as a single peak of increased read depth since they are present twice in the genome but usually only once in the short read assembly (Li *et al.*, 2014; Merrill *et al.*, 2016; Garneau *et al.*, 2017).

Headful packaging and circularly permuted genomes

The genomes of phages in the family *Straboviridae* and *Drexlerviridae* do not have defined ends like the *Autographiviridae*, and differ in that they are circularly permuted and DNA is packaged by the headful method (**Figure 1.5C**) (Kutter *et al.*, 2005; Casjens and Gilcrease, 2009; Merrill *et al.*, 2016). Circular permutation is the arrangement of a linear piece of DNA which is terminally redundant and the sequence can be arranged in a circle when a population of genomes is taken together (Kutter *et al.*, 2005). That is, sequences in individual phage capsids can have the arrangement ABCDEA, BCDEAB, CDEABC, DEABCD, and so on.

A specific DNA sequence known as the *pac* site dictates the location of only the first cleavage of the concatemer in some phages with circularly permuted genomes (Casjens and Gilcrease, 2009; Merrill *et al.*, 2016; Weber-Dąbrowska *et al.*, 2023). However, the *pac* site has not been determined for all phages that utilise headful packaging, such as the model phage T4, and it appears that packaging initiation starts at a random site (Alam *et al.*, 2008; Garneau *et al.*, 2017). Subsequent cleavage sites by the terminase protein are different for every concatemer: as the DNA is packaged, the next concatemer is cut only when capsid is full (hence headful) (Casjens and Gilcrease, 2009; Li *et al.*, 2014; Merrill *et al.*, 2016). Each capsid will thus contain 102-110% of the genome with different beginning and end (i.e. repeat) sequences to one another, and different lengths (Casjens and Gilcrease, 2009; Li *et al.*, 2014; Merrill *et al.*, 2016). Practically, sequencing of a circularly permuted genome can result in an assembly with more than 100% of the genome length since there are no clearly defined "ends" (Merrill *et al.*, 2016). Assemblies that are significantly longer than 100% have been recently shown to occur more often when using long-read sequencing for these types of phages (Cook *et al.*, 2024).

1.3.3 The *Drexlerviridae* (T1-like) phages

The *Drexlerviridae* are a family of phages comprised of four subfamilies, 25 genera, and 116 species (Walker *et al.*, 2021; ICTV, 2023; Turner *et al.*, 2023). The subfamily *Tunavirinae* is comprised of three genera, with the eponymous *Escherichia* phage T1 assigned to the genus *Tunavirus* (ICTV, 2023). The genus *Webervirus* has not been assigned a subfamily however, and currently contains 32 species and many more unclassified species, almost all of which target *Klebsiella* spp. (ICTV, 2023).
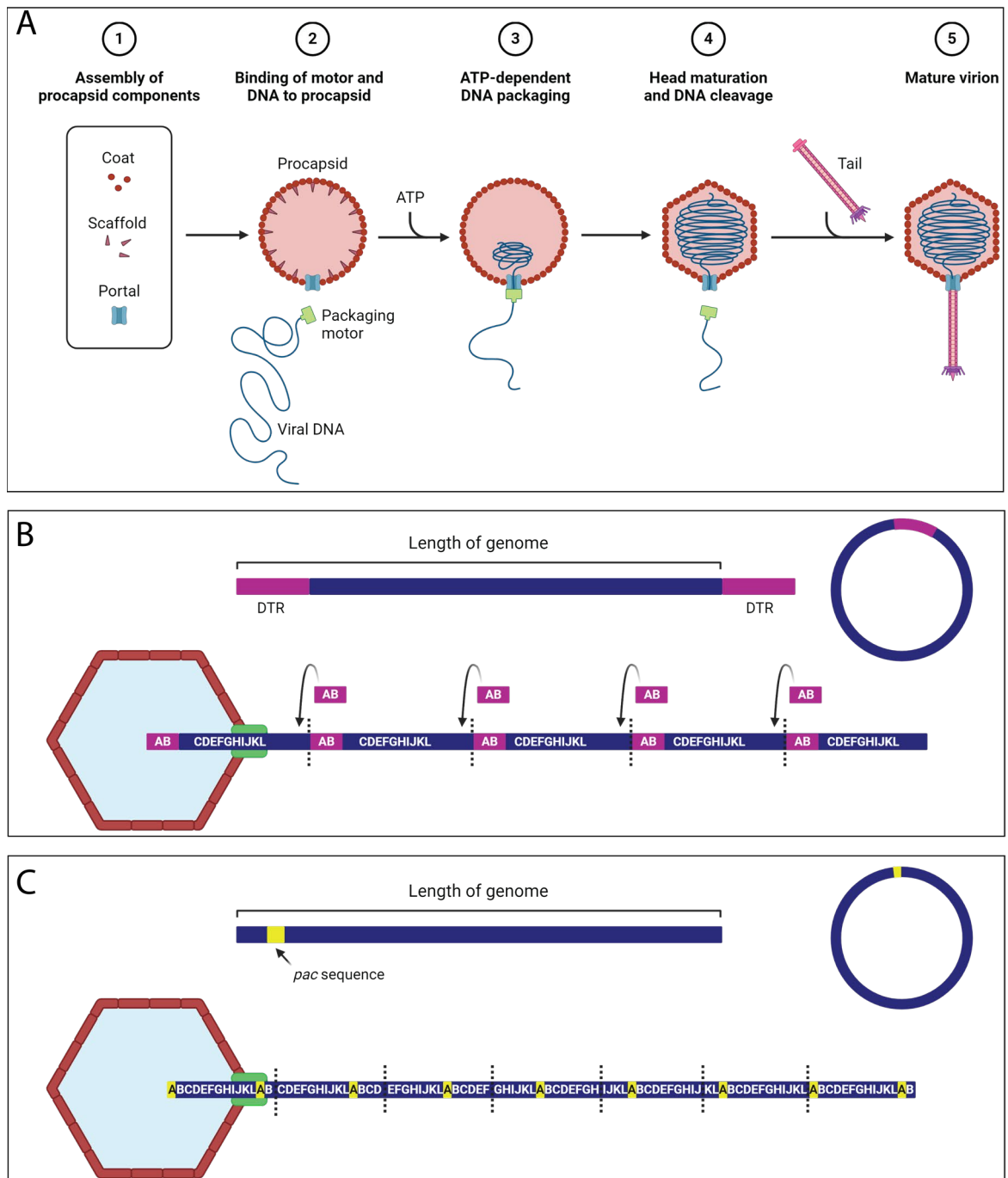
Figure 1.5. Schematic representation of phage DNA packaging methods to generate distinct termini. (A) Generalised schematic of the DNA packaging process in tailed phages catalysed by the terminase (packaging motor). **(B)** Packaging method of genomes with DTRs. The phage genome consists of linear dsDNA flanked by short DTR (top panel), which then circularises upon host entry (right panel). Rolling circle replication generates a linear concatemer of DNA with a single repeat at one end, and a second repeat is synthesised during packaging into the phage capsid (bottom panel). Dotted lines represent terminase cut sites. **(C)** Headful packaging method that generates terminally redundant, circularly permuted genomes. The genome consists of *pac* sequences (yellow) (top panel), which also circularises upon host entry (right panel). Rolling circle replication generates a linear concatemer of DNA with a *pac* sequence that dictates the terminase cut site, usually for the first cleavage only (dotted line). Each concatemer is cut at a slightly different site when the head is full so that each linear piece of DNA contains more than 100% of the genome (bottom panel). The DNA is packaged into the capsid, which will then contain multiple copies

22

of the genome with different starting sequences. Figure generated in BioRender and adapted from Moore and Prevelige (2002); Merrill *et al.* (2016); Hilbert *et al.* (2017).

Phages belonging to the family *Drexlerviridae* have genomes that are circularly permuted and terminally redundant (Guttman *et al.*, 2005; Khan *et al.*, 2021), as described above. Some phages in the family *Drexlerviridae* encode one or two tRNAs (Khan *et al.*, 2021), which are thought to assist phages in continuation of host takeover as the host cellular machinery is degraded due to phage infection (Bailly-Bechet *et al.*, 2007; Yang *et al.*, 2021). Typically, their genomes are approximately 43-52 kbp and genes are bidirectional, i.e. they encode proteins on both the forward and reverse strands.

Drexlerviridae phages likely target specific O-antigen moieties and porins as their primary and secondary receptors, respectively (Maffei *et al.*, 2021). Specifically, *Escherichia* phage T1 targets FhuA, while other T1-like phages were shown to target BtuB, LptD, FepA, YncD, and TolC (Kutter *et al.*, 2005; Maffei *et al.*, 2021; Acton *et al.*, 2024). FhuA, FepA, YncD are membrane-bound proteins involved in iron acquisition (Kutter *et al.*, 2005; Hesse *et al.*, 2020; Maffei *et al.*, 2021; Federici *et al.*, 2022), and BtuB a vitamin B12 transporter (Pieńko and Trylska, 2020; Maffei *et al.*, 2021; Acton *et al.*, 2024). LptD is involved in LPS transport and assembly (Janssen *et al.*, 2020), whereas TolC is an important component of the AcrAB-TolC efflux pump (Zgurskaya *et al.*, 2011; Blair *et al.*, 2015).

Some studies (Ku *et al.*, 2021) performed with *Klebsiella* phages in the family *Drexlerviridae* have indeed focused on phenotypic analyses in the context of phage therapy, but are lacking in genomic and functional genomic work to characterise phages.

1.3.4 The *Straboviridae* (T4-like) phages

Within the family *Straboviridae*, there are currently three subfamilies, 36 genera, and 208 species (ICTV, 2023; Turner *et al.*, 2023). The subfamily *Tevenvirinae* comprises 14 genera, including *Jiaodavirus* and *Tequatrovirus*, the latter containing the eponymous *Escherichia* phage T4 (ICTV, 2023). The genus *Slopekvirus* has not been assigned to a subfamily and its members are more distantly related to T4 than the genus *Jiaodavirus* (ICTV, 2023; Turner *et al.*, 2023).

Phages belonging to the family *Straboviridae* have highly conserved genomes that are circularly permuted and terminally redundant (Kutter *et al.*, 2005; Smith-Zaitlik *et al.*, 2022), as described above. They have characteristic prolate capsids and their genomes are large, at more than 160 kbp (Guttman *et al.*, 2005). Genes are encoded bidirectionally and densely packed, with many protein coding sequences overlapping (Brüssow and Kutter, 2005). They also encode multiple tRNAs (van den Berg *et al.*, 2023).

T4-like phages are thought to target porins such as OmpA, OmpK35 (OmpF homolog), or OmpK36 (OmpC homolog) as their primary receptors (Guttman *et al.*, 2005; Kortright *et al.*,

2020; Maffei *et al.*, 2021). Interestingly, T4-like phages are highly adaptable and capable of targeting alternative primary receptors such as LPS components (Guttman *et al.*, 2005; Kortright *et al.*, 2020; Maffei *et al.*, 2021). LPS components have also been identified as secondary receptors in T4 (Kutter *et al.*, 2005). Other primary receptors include Tsx, which is involved in deoxynucleoside transport across the OM (Nieweg and Bremer, 1997; Kortright *et al.*, 2020; Maffei *et al.*, 2021). LamB and FadL have also been implicated as receptors in functional screens of *E. coli* T4-like phages, which function as selective transporters of specific sugars and fatty acids, respectively (Kortright *et al.*, 2020; Maffei *et al.*, 2021). Federici *et al.* (2022) found potential receptors targeted by an unclassified T4-like phage in the subfamily *Tevenvirinae* included LPS components (*rfaQ*), stress response factors (*grcA*), and type II secretion system components (*outN* also called *pulN*) (Kwon *et al.*, 2016; Arizala and Arif, 2019; Chernyatina and Low, 2019; Yang *et al.*, 2019). Moreover, a second T4-like phage in the genus *Slopekvirus*, found iron acquisition genes (*fhuA*) as potential phage receptors (Federici *et al.*, 2022), and has been reported previously (Hesse *et al.*, 2020).

1.3.5 The *Autographiviridae* (T7-like) phages

Within the family *Autographiviridae*, there are nine subfamilies, 134 genera, and 376 species, representing one of the largest groups of classified phages thus far. The subfamily *Studiervirinae* comprises 24 genera, of which the genus *Przondovirus* contains 29 species.

The genome organisation of genera within the family *Autographiviridae* is highly conserved: almost all genes are unidirectional and show a high degree of synteny, and linear genomes are flanked by DTRs (Molineux, 2006; Adriaenssens *et al.*, 2011; Adriaenssens *et al.*, 2020; Evseev *et al.*, 2020; Boeckman *et al.*, 2022), as previously described. Their genomes are approximately 38-42 kbp.

Genes are clustered into early, middle, and late genes responsible for host takeover, phage genome replication, and virion production followed by host cell lysis, respectively (Guttman *et al.*, 2005; Severinova and Severinov, 2006; Clokie *et al.*, 2011; Evseev *et al.*, 2020). In T7, transcription occurs as the genome enters the cytoplasm, with early genes expressed by the host *E. coli* RNA polymerase (RNAP), including proteins that inactivate host-mediated transcription (Guttman *et al.*, 2005; Molineux, 2006; Severinova and Severinov, 2006). Then, T7-encoded DNA-directed RNAP is expressed, which takes over transcription of middle and late genes (Guttman *et al.*, 2005; Molineux, 2006; Severinova and Severinov, 2006; Savalia *et al.*, 2010).

Phage T7 and many T7-like phages rapidly lyse their bacterial hosts, completing their life cycle in less than 20 minutes (Molineux, 2006).

One of the first genes to be expressed in T7-like viruses is the Ocr protein, or S-adenosyl-L-methionine hydrolase (T7_gp0.3 or K30_gp0.3, respectively), which are both involved in

counteracting host type I restriction-modification (RM) systems (Maffei *et al.*, 2021). Both genes are good markers for the start of the genome and determining the approximate location of the DTRs.

A unique feature of T7-like phage infection is the release of three internal virion proteins (T7_gp14, T7_gp15, and T7_gp16) that insert into the host cell envelope, creating a pore-like structure that facilitates DNA entry (Molineux, 2006; González-García *et al.*, 2015; Nobrega *et al.*, 2018; Maffei *et al.*, 2021). Receptor binding then leads to a conformational change in the short, non-contractile tail tube, and DNA is ejected into the host (Molineux, 2006; González-García *et al.*, 2015; Nobrega *et al.*, 2018; Maffei *et al.*, 2021).

To overcome the reliance on host-mediated transcription, phages in the family *Autographiviridae* encode their own DNA-directed RNAP, a hallmark of the group (Molineux, 2006; Savalia *et al.*, 2010; Choi, 2012). Interestingly, only a very small proportion of the phage DNA is injected into the host cell at any given time and is transcribed *de novo* by their own RNAP (Savalia *et al.*, 2010; Choi, 2012). The T7-like phage RNAPs exhibit high specificity for their respective promoter sequences, and this feature has enabled T7 RNAPs and their promoters to be widely utilised in molecular biology (Studier and Moffatt, 1986; Guttman *et al.*, 2005; Molineux, 2006).

LPS components are almost always identified as the secondary receptor in Gram-negative-targeting autographiviruses, but distinguishing between primary and secondary receptors and the role of the capsule is yet to be fully elucidated (Molineux, 2006; Maffei *et al.*, 2021).

1.4 Bacteriophage therapy

1.4.1 Early phage therapy

Phages were discovered independently in the early 1900s by Frederick Twort and Felix d'Hérelle after observing bacterial lysis (Twort, 1915; d'Hérelle, 1917). d'Hérelle realised the potential to utilise phages as therapeutics to control bacterial infections, termed bacteriophage (phage) therapy. Having pioneered early phage therapy, d'Hérelle demonstrated efficacy in patients with bacterial dysentery and in *Salmonella*-infected chickens (d'Hérelle, 1917; d'Hérelle and Smith, 1926). The absence of phage characterisation however, further compounded by gaps in the knowledge of phage biology and the causal agents of disease, led to poorly designed studies that contradicted d'Hérelle's prior success with phage therapy (Kortright *et al.*, 2019). With the discovery of antibiotics, these factors led to the abandonment of phage therapy in Western medicine. Its development had continued in the former Soviet Union however, and is still practiced in Russia and Georgia today (Kortright *et al.*, 2019).

1.4.2 Key advantages of phage therapy

Limited microbiome side effects. An advantage of the use of phages over antibiotics in treating bacterial infections is their high specificity due to a narrow host range, which limits perturbations on the human microbiome. Antibiotic therapy disrupts the gut microbiota and selects for resistant pathogens. Moreover, eradication of protective species can result in secondary infections including antibiotic-associated diarrhoea (Högenauer *et al.*, 2006; Sequeira *et al.*, 2020). However, a recent study in a murine model demonstrated that phage therapy does indirectly alter the bacterial composition of the gut microbiome (Hsu *et al.*, 2019). Indeed, predation of the host bacteria by a two-phage cocktail led to enrichment of some non-susceptible species and reduction of others (Hsu *et al.*, 2019). Whilst this study provides ecological insights into phage predation within the gut microbiome, the implication for phage therapy to indirectly modulate bacterial composition is yet to be fully elucidated. A recent clinical trial found that phages targeting *K. pneumoniae* in healthy volunteers exhibited no adverse effects on the gut microbiome (Federici *et al.*, 2022).

Self-replicating and self-limiting. Exponential replication of lytic phages continues in the presence of the host bacteria at the site of infection, followed by rapid clearance as the host is eradicated. This self-replicating, self-limiting nature of lytic phages also make them an attractive alternative to antibiotics (Romero-Calle *et al.*, 2019).

Mitigation of resistance development. Phage therapy can also overcome some intrinsic resistance mechanisms among bacteria that typically decrease antibiotic efficacy. These include disruption of biofilms and capsule degradation by phage-encoded enzymes. For example, phage TSK1 demonstrated anti-biofilm activity against an MDR *K. pneumoniae* strain *in vitro* (Tabassum *et al.*, 2018). The ability of some phages to express capsular depolymerases as part of their tail structure allows phages to access the bacterial cell for infection to proceed (Pan *et al.*, 2015; Hsieh *et al.*, 2017; Pan *et al.*, 2017; Cai *et al.*, 2019). Moreover, purification of a capsule depolymerase have shown to increase the efficacy of subsequent antibiotic therapy (Bansal *et al.*, 2014). Together, these demonstrate important mechanisms of phages to overcome intrinsic resistance among *Klebsiella* spp.

Whilst it can be argued that bacterial resistance to phage infection is a limitation, this can also be advantageous as phages can co-evolve to overcome this resistance (Pirnay *et al.*, 2018). Moreover, phage resistance may also result in a concomitant reduction in bacterial virulence and/or antibiotic resistance (Hesse *et al.*, 2020). For example, OmpK36 was shown to be a key receptor for phage GH-K3 infection, and the phage-resistant Δ*ompK36* mutant appeared to show defective iron uptake (Cai *et al.*, 2018). Indeed, separate studies have previously demonstrated porin mutants are attenuated *in vivo* (Tsai *et al.*, 2011; Wong *et al.*, 2019). These outcomes may reduce infection severity to allow more efficient immune clearance, or re-sensitise the bacteria to antibiotics. Another recent study demonstrated that resistance development in *Salmonella enterica* subsp. *enterica* serovar Typhimurium

against one phage resulted in re-sensitisation of another, suggesting that combination of the two phages can increase efficacy of combinations by inducing collateral sensitivity (Acton *et al.*, 2024).

Phage-antibiotic synergy is a phenomenon whereby the combined effects of phage and sub-inhibitory concentrations of an antibiotic is greater than their sum (Comeau *et al.*, 2007). Phage-antibiotic synergy is advantageous since it can be exploited to not only increase the efficacy of either treatment, but also to reduce the selection for resistance determinants (Abedon, 2019; Tagliaferri *et al.*, 2019), and has been reported with Gram-negative bacteria (Comeau *et al.*, 2007; Jansen *et al.*, 2018; Lin *et al.*, 2018).

Prophylactic use. Phage therapy can also be used as a preventative measure. For example, Townsend *et al.* (2020) found that the application of a *Klebsiella* phage cocktail significantly prevented biofilm formation, thereby reducing infection risk in a catheter model. Moreover, phage-antibiotic synergy was highly efficacious when compared to antibiotic treatment alone (Townsend *et al.*, 2020). Federici *et al.* (2022) also found administration of a cocktail of five phages led to very low *K. pneumoniae* abundance in the gut of healthy volunteers, highlighting the potential as a prophylactic in preventing the development of IBD in healthy individuals. Moreover, prophylactic application *in vitro* resulted in inhibition of biofilm formation (Tabassum *et al.*, 2018). Prophylactic use may be of high clinical value in preventing *Klebsiella* biofilm formation on medical devices such as catheters (Chhibber *et al.*, 2013). Kelly and Jameson (2023) were recently able to show that prophylactic treatment with a phage cocktail successfully rescued *Galleria mellonella* following infection with *K. pneumoniae*.

1.4.3 Key limitations of phage therapy

Phage resistance. Phage resistance, like that for antibiotics is an evolutionary phenomenon. One of the key regulatory hurdles to widespread adoption of phage therapy in the UK is the development of resistance to phages (described below). Previous work has shown that bacteria rapidly develop resistance to phages, usually through receptor modification (Chevallereau *et al.*, 2022). For example, phages that utilise porins as a primary or secondary receptor would select for porin deletion mutants (Cai *et al.*, 2018; Hesse *et al.*, 2020). This phenotype could reduce both phage susceptibility as well as reduce antibiotic influx, thereby leading to cross-resistance to both phage and antibiotics. The modification of OM-bound components is a short-term consequence of co-evolution, whereas the development of more complex resistance mechanisms such as anti-phage defence systems is long-term consequence of co-evolution (Chevallereau *et al.*, 2022)

Narrow host range. Whilst the narrow host spectrum of phages is advantageous, it can also be a limiting factor to their therapeutic use. Not only can phage cocktails reduce the emergence of phage resistance, they can also further increase host range (Romero-Calle

et al., 2019). The combination of *Klebsiella* phages expressing distinct depolymerases would be advantageous for targeting highly diverse capsular serotypes, for example. However, intimate knowledge of interactions within a cocktail is necessary to mitigate any potential antagonism among these phages that may reduce stability and efficacy (Hsu *et al.*, 2019; Kortright *et al.*, 2019; Hesse *et al.*, 2020).

The right phage for the job. *In vitro* characterisation of phages is important to inform *in vivo* studies and assess their clinical value (Abedon, 2017). Temperate phages are not suitable for phage therapy as they do not lyse the bacterial cell immediately, making them less efficacious than lytic phages (Romero-Calle *et al.*, 2019). Additionally, temperate phages can transfer antibiotic resistance determinants, virulence factors, or toxin genes to the host in a process called lysogenic conversion (Hay and Lithgow, 2019), presenting major safety concerns. An example of such is the cholera toxin that is introduced to *Vibrio cholerae* by the temperate inovirus CTXphi (Waldor and Mekalanos, 1996; Hay and Lithgow, 2019). For use as therapeutic agents, phages and their life cycle need to be well-characterised, including identification of genes associated with lysogeny, and any type of virulence factor or gene that can confer an advantage to the host (Abedon, 2017), emphasising the need for complete and accurate phage genomes.

Immune response. Trans-kingdom interactions may also influence phage therapy acceptance into Western medicine. Potential immune stimulation by phages may have several consequences, including the immune clearance of phage and overstimulation of the immune response (Romero-Calle *et al.*, 2019). Immune clearance by both the innate and adaptive immune responses have been reported: one study (Geier *et al.*, 1973) demonstrated rapid elimination of phage in germ-free mice by macrophages, while another (Hodyra-Stefaniak *et al.*, 2015) demonstrated phagocytic clearance and antibody neutralisation of phage in a murine model of systemic inflammation. Łusiak-Szelachowska *et al.* (2014) however, found that antibody neutralisation of phage in healthy volunteers was dependent on phage type and administration route. Anti-phage antibodies were also reported within the gut mucosa of a murine model after oral administration of T4 phage therapy (Majewska *et al.*, 2015). Whilst efficacy of the phage preparation was reduced, immunogenicity was low and the study required sustained and high-titre treatment to produce an immune response (Majewska *et al.*, 2015). Another study (Lin *et al.*, 2014) reported that persistence of phage following therapy in a murine model of HvKP failed to elicit an immune response.

Production of high-quality phage preparations. Endotoxin, the lipid A moiety of LPS, is anchored to the bacterial OM and is released upon lysis (Lepper *et al.*, 2002). Large amounts can be released during bacterial replication in addition to lysis, which can form large, highly stable aggregates (Gorbet and Sefton, 2005). Moreover, endotoxin aggregates carry a net negative charge, binding easily to cations (Gorbet and Sefton, 2005). Endotoxin

is a common contaminant in laboratory and industry settings, and none more so than in phage preparation settings (Gorbet and Sefton, 2005). Endotoxin is highly immunogenic; a contributing factor to overstimulation of the immune response and poor outcomes during infection, particularly in cases of sepsis (Lepper *et al.*, 2002; Dufour *et al.*, 2019). Regulatory approval centres around the safety implications associated with residual bacterial endotoxin-contaminating phage preparations (Pelfrene *et al.*, 2016), or rapid release of endotoxin following phage-induced lysis (Pelfrene *et al.*, 2016; Jault *et al.*, 2019). However, Dufour *et al.* (2017) found that phage therapy released less endotoxin when compared to β-lactam antibiotics, which are also rapidly bactericidal. Moreover, the same group reported that neither phage-induced endotoxin release, nor virus particles themselves resulted in overstimulation of the immune response during phage therapy when compared to β-lactam antibiotics (Dufour *et al.*, 2019). Taken together, these studies indicate immune responses to phage therapy may have efficacy and safety implications, factors of which may be phage-specific and/or study design-specific (Dufour *et al.*, 2019).

An example of a trial where the manufactured phage preparation was not extensively characterised for its stability, was the burn wound trial Phagoburn. Here, a good manufacturing practice (GMP)-compliant 12-phage cocktail in burn wound patients suffered a multitude of problems, including reduced titre (Jault *et al.*, 2019). Each phage was stored separately and only combined prior to treatment, which led to a significant reduction in phage stability and thus, titre and efficacy (Jault *et al.*, 2019).

Phage-antibiotic combinations can also be antagonistic. Whilst phage-antibiotic synergy is advantageous, several studies have reported either antagonistic effects or no effect at all, even at higher concentrations of antibiotic (Verma *et al.*, 2009; Abedon, 2019). It is difficult to predict which phage-antibiotic combinations would be most useful in treating infections caused by *Klebsiella* spp. (Abedon, 2019), thus further investigation into the molecular mechanisms underpinning phage-antibiotic synergy are necessary.

Regulatory hurdles. The emergence of phage resistance is a compounding problem to the adoption of phage therapy. The current regulatory framework within the EU places phage therapy under the biological medicinal products directive (European Commission, 2001), which constrains the adaptation of phages or phage cocktails to deliver personalised treatments in a timely and cost-effective manner (Furfaro *et al.*, 2018; Pirnay *et al.*, 2018). This emphasises the need for an independent framework supporting the use of biologically-active products that can be used in personalised medicine (Pirnay *et al.*, 2018). Phage resistance from a regulatory perspective can be addressed through the production of a fixed phage cocktail, by combining two or more phages within a single preparation (Pelfrene *et al.*, 2016; Romero-Calle *et al.*, 2019; Hesse *et al.*, 2020). Alternatively, individual pre-approved phages can be selected from a phage bank as appropriate (Pelfrene *et al.*, 2016). In Belgium, the use of phage therapy as magistral preparations has allowed clinicians to

circumvent the current regulatory hurdles (Pirnay *et al.*, 2018). This allows the prescribing and preparation of high-purity phage products in line with patient needs, and provides the necessary flexibility to rapidly adapt phage products for personalised therapy (Pirnay *et al.*, 2018; Romero-Calle *et al.*, 2019). A considerable bottleneck in the adoption of phage therapy in the UK is a lack of standardised methodology for the purification of phages (Larsen *et al.*, 2023; Suleman *et al.*, 2024). The use of phages in compassionate cases i.e. those for whom all other treatment options have failed in the UK has previously relied on sourcing phages from the US and/or Belgium that have produced non-GMP high-purity, high-titre phage preparations (Dedrick *et al.*, 2019; Jones *et al.*, 2023; Suleman *et al.*, 2024). Ironically however, regulatory approval of any therapeutic in the UK requires production to GMP, which ensures that medicinal products are quality controlled to a particular set of internationally recognised standards (Jones *et al.*, 2023; Suleman *et al.*, 2024).

1.5 A short overview of *Klebsiella*-targeting bacteriophages

Many *Klebsiella* phages described so far have had to adapt to the diversity of CPS among *Klebsiella* spp. (Hsieh *et al.*, 2017; Latka *et al.*, 2019). A recent meta-analysis (Latka *et al.*, 2019) of *Klebsiella* phage genomes demonstrated that *K. pneumoniae* phage RBPs are also highly diverse. Indeed, rapid modifications to their structural and genetic composition allow the phages to expand their host range by targeting multiple capsular serotypes (Latka *et al.*, 2019). Moreover, *Klebsiella* phages with diverse RBPs can express a range of different depolymerases capable of degrading multiple capsular serotypes (Hsieh *et al.*, 2017; Pan *et al.*, 2017; Cai *et al.*, 2019; Latka *et al.*, 2019). For example, *K. pneumoniae* phage phiK64-1, was found to express eight novel depolymerases that contributed to its expanded host range (Pan *et al.*, 2015; Pan *et al.*, 2017). Additionally, K5-2 phage was capable of infecting seven different capsular types of *K. pneumoniae*, which was owed to the presence of two RBPs each with depolymerase activity (Hsieh *et al.*, 2017).

Some *Klebsiella* phages appeared to be limited to a specific capsular type however, demonstrating a narrow host range. For example, Gao *et al.* (2022), described a novel phage that could lyse three *K. pneumoniae* strains with the capsular type K63, but not other capsular types. Moreover, Haudiquet *et al.* (2021) found that phage predation drives loss of capsular components, but that subsequent acquisition of conjugative plasmids was higher among the acapsular mutants, facilitating reacquisition of capsule determinants. Thus, not only do phages drive serotype swap through the reacquisition of capsular components, conjugation may also drive the concomitant acquisition of AMR determinants (Haudiquet *et al.*, 2021). Interestingly, Lourenço *et al.* (2023), found that phages targeting acapsular mutants of *K. pneumoniae* demonstrated a broader host range and that resistance development was significantly reduced. Likely explanations could be that the capsule is often lost under specific conditions, and phages that are capsule-reliant can also drive this loss (Buffet *et al.*, 2021; Haudiquet *et al.*, 2021; Lourenço *et al.*, 2023). Moreover, phages

targeting acapsular mutants tend to target more highly conserved bacterial components than their capsule-reliant counterparts, resulting in reduced resistance development due to the associated fitness cost (Lourenço *et al.*, 2023). However, just as phages can drive capsular diversity, capsular diversity can also drive phage RBP and depolymerase diversification to increase their host range (Beamud *et al.*, 2023).

Few studies have focused on identification of receptors for both reversible and irreversible attachment among *Klebsiella* phages. However, modelling phage resistance can give insights into phage RBPs and their targeted receptors. Resistance to phage infection can occur through receptor loss or downregulation, as well as through inability to access the receptor by capsule or biofilm production, for example (Hsieh *et al.*, 2017). Phage-resistant mutants of *K. pneumoniae* have shown reduced virulence through inhibited LPS synthesis, capsule synthesis, and reduced fimbrial expression, suggesting these components act as phage receptors (Tan *et al.*, 2019; Hesse *et al.*, 2020). Lin *et al.* (2014) found that infection of *K. pneumoniae* by phage NTUH-K2044-K1-1 could not occur in acapsular mutants, suggesting CPS as the primary phage receptor. Tomás *et al.* (1987) suggested that the O-antigen of LPS was a phage receptor for three phages, FC3-2, FC3-3, and FC3-6, as phage resistance in *K. pneumoniae* was mediated by O-antigen loss. Another study (Cai *et al.*, 2018) demonstrated that OmpK36 was essential to phage GH-K3 infection in *K. pneumoniae* through the analysis of OmpK36 deletion mutants (Δ*ompK36*), and most likely the targeted receptor for irreversible attachment. These studies help to inform phage cocktail design by selecting phages with different host receptors, that make the selection of resistance less likely following therapeutic application (Hesse *et al.*, 2020).

1.5.1 Brief introduction into phage therapy case studies for *Klebsiella* spp.

With the emergence and increasing prevalence of antibiotic resistance among pathogenic bacteria, novel approaches to tackle the current crisis are imperative. Previous efforts have focused on structural modification of existing antibiotics or the development of inhibitors of resistance mechanisms (Blair *et al.*, 2015; Theuretzbacher, 2017; Malmir *et al.*, 2018). This reactive approach is of limited clinical value, since increased efficacy is often short-lived (Theuretzbacher, 2017). Phage therapy has thus seen a resurgence in recent years as an alternative or adjunctive to current antibiotic therapy, particularly for recalcitrant infections caused by priority pathogens such as *Klebsiella* spp.

Recently, pre-adapted phages were used in conjunction with antibiotics to treat a persistent PDR *K. pneumoniae* infection in a bombing survivor (Eskenazi *et al.*, 2022). The patient had received extensive antibiotic treatment over a prolonged period of time, including the nephrotoxic antibiotic colistin (Eskenazi *et al.*, 2022). Pre-adaptation of the phage, a member of the genus *Slopekvirus*, was performed by co-evolution with the patient's strain

of *K. pneumoniae*, and when combined with β-lactam and polymyxin antibiotics, led to complete clearing of the infection three years post-treatment (Eskenazi *et al.*, 2022).

Another study found that a cocktail of five phages could be used as a prophylactic in healthy volunteers for the prevention of inflammatory bowel disease (IBD) development by removal of gut commensal *K. pneumoniae* (Federici *et al.*, 2022). Moreover, the phage combination suppressed *K. pneumoniae*-mediated inflammation in an IBD animal model (Federici *et al.*, 2022). The five phages spanned three families (*Straboviridae*, *Autographiviridae*, and *Demerecviridae*), and whilst some displayed differences in their predicted host receptor, there was some overlap (Federici *et al.*, 2022).

Phage therapy was also shown to be successful in resolving *K. pneumoniae* infection in the prosthetic joint of a patient, in whom amputation had been recommended (Cano *et al.*, 2021). A single-phage GMP-grade preparation was administered to the patient, not only resolving the infection, but preventing re-infection by the same organism up to 34 weeks post-treatment, negating the need for amputation of the limb (Cano *et al.*, 2021).

Recently, Kelly and Jameson (2023) were able to show a phage cocktail could successfully rescue a *Klebsiella* infection model when used as both a prophylactic and as an active treatment. Moreover, the phage cocktail was designed to encompass as broad host range as possible, including phages from distinct genetic lineages capable of infecting six distinct capsular types of *Klebsiella* spp., and was produced with low endotoxin levels (Kelly and Jameson, 2023).

1.6 Transposon mutagenesis

Transposon mutagenesis involves the use of a transposon to disrupt gene coding sequences and/or regulatory sequences to determine the phenotype and/or expression of the gene in question (Hoffman *et al.*, 2000; Cain *et al.*, 2020). Transposons are small sequences of DNA capable of moving within or between genomes, and are used in combination with a transposase enzyme to facilitate genomic insertion (Hoffman *et al.*, 2000; Reznikoff, 2008; van Opijnen and Camilli, 2013; Bourque *et al.*, 2018). The transposase recognises sequence-specific sites within the transposon DNA, excising the transposon and inserting it into the target genome, which can be site-specific or randomly inserted, depending on the transposon type (Hoffman *et al.*, 2000; van Opijnen and Camilli, 2013; Cain *et al.*, 2020). One of the most commonly used in molecular biology applications is the Tn5 transposon and its transposase coupled to a selectable marker such as an antibiotic resistance cassette (Hoffman *et al.*, 2000; Reznikoff, 2008).

Transposon mutagenesis can be used to explore the genotype-phenotype link, something that is considerably lacking within phage functional studies (Holden *et al.*, 2021; Maffei *et al.*, 2021). Moreover, the combination of transposon mutagenesis with NGS, generally

referred to as transposon-insertion sequencing (TIS), can be used as a high-throughput functional screen of genes and regulatory sequences involved under specific conditions (Langridge *et al.*, 2009; Jana *et al.*, 2017; Cain *et al.*, 2020; Holden *et al.*, 2021; Acton *et al.*, 2024).

Transposon-directed insertion site sequencing (TraDIS) is a specific type of TIS that relies on the creation of a dense mutant library through large-scale transposon mutagenesis, comprising of approximately 400,000 mutants, but can be as large as 1 million mutants (Langridge *et al.*, 2009; Holden *et al.*, 2021). The library is composed of single gene mutants that are pooled, allowing high-throughput assay of multiple genes simultaneously under various experimental conditions (Langridge *et al.*, 2009; Holden *et al.*, 2021). Specifically, the transposon mutant library can be used to determine genes involved in the bacterial response to antibiotic or phage stress, for example (Turner *et al.*, 2020; Holden *et al.*, 2021; Acton *et al.*, 2024). This is achieved through sequencing of the transposon inserts to identify their location and comparing them to a non-stressed control (Cain *et al.*, 2020; Turner *et al.*, 2020; Holden *et al.*, 2021).

A general overview of TraDIS involves i) mutant library construction and pooling; ii) DNA extraction from the mutant pool, fragmentation of the DNA, addition of adapters, and amplification and purification of DNA fragments containing the transposon; iii) sequencing of the transposon-enriched fragments and mapping them back to the genome (**Figure 1.7A**). This process determines the number of mutants within the mutant library pool. Next, the mutant library is experimentally challenged with antibiotics or phage infection, for example, alongside a non-stress control (**Figure 1.7B**). The mutant library DNA under both the stress and control conditions is extracted and processed in the same way as described in steps ii) and iii). Next, the number and location of the transposon insertions are directly compared between the two conditions, and can be used to determine the relative fitness of specific genes under those conditions tested (**Figure 1.7B**). For example, increased insertional frequency in a particular gene relative to the control would suggest mutants have increased fitness under these conditions (Cain *et al.*, 2020; Acton *et al.*, 2024). Conversely, reduced frequency of insertions in a particular gene relative to the control would suggest mutants have reduced fitness under the conditions (Cain *et al.*, 2020; Acton *et al.*, 2024).

TraDIS can been used to understand genes involved in survival, stress responses, antimicrobial tolerance, virulence, and even phage infection in a range of pathogenic bacteria (Langridge *et al.*, 2009; Jana *et al.*, 2017; Yasir *et al.*, 2020; Bruchmann *et al.*, 2021; Holden *et al.*, 2021; Acton *et al.*, 2024). In *Klebsiella* specifically, TraDIS has been used to understand infection dynamics, virulence, and the AMR of *K. pneumoniae* (Jana *et al.*, 2017; Short *et al.*, 2020; Bruchmann *et al.*, 2021).
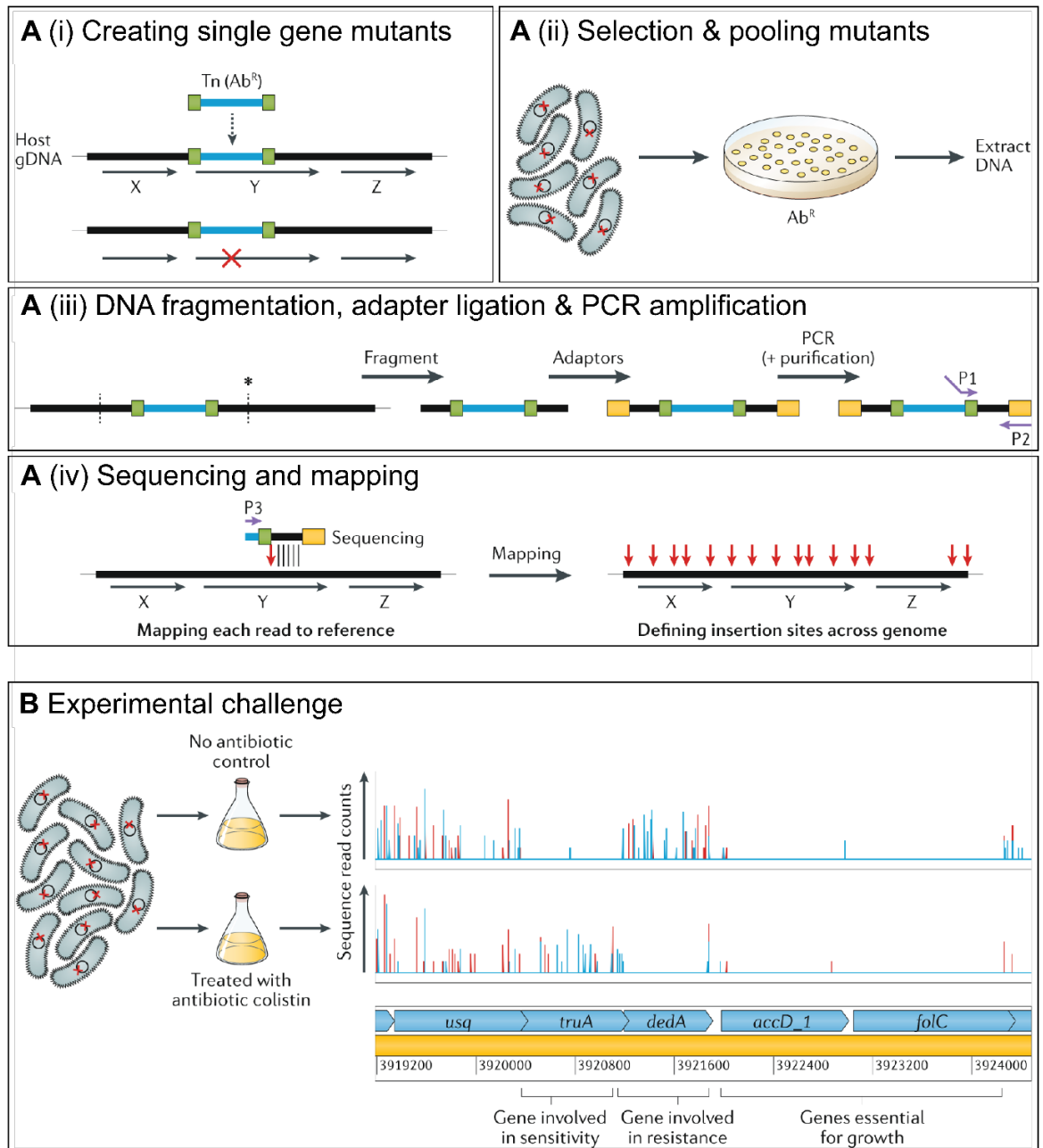
Figure 1.6. Overview of TraDIS mutant library construction, sequencing, and experimental challenge. A (i) Creation of single gene mutant library through transposon (Tn, blue) insertion into the host genome (gDNA, black) with a selectable marker, usually an antibiotic resistance cassette (Ab^R). Insertion within a particular gene results in its disruption (red cross). **A (ii)** Once the mutant library has been created, mutants are plated with the selectable marker and pooled together, and the DNA extracted from the pooled sample. **A (iii)** DNA is fragmented, and adapters ligated to fragments to PCR amplify and purify denoted by primers (P1 and P2), so only fragments containing the transposon are enriched. **A (iv)** DNA fragments containing the transposon are sequenced using primers (P3) and reads mapped to the bacterial genome to determine the number and location of each transposon mutant (red arrows). Genes without any transposon insertions are determined as essential genes. **(B)** Experimental challenge of the transposon mutant library with, for example, the antibiotic colistin (bottom panel) and directly compared to the control group (top panel). Insertion frequency is given by the height of the lines and colour denotes onto which strand the transposon has been inserted (red and blue for forward and reverse, respectively). Higher frequency of insertions in the control relative to the challenged sample

34

(e.g. *truA*) denote mutants with increased fitness under the conditions (i.e. a susceptibility gene). Reduced frequency of insertions in the control relative to the challenged sample (e.g. *dedA*) denote mutants with reduced fitness under the conditions (i.e. a resistance gene). Genes without insertions are determined as essential genes (e.g. *accD_1* and *folC*). Figure adapted from Cain *et al.* (2020).

1.7 Aims and objectives

The aim of this project was to develop a strategy for the biocontrol of *Klebsiella* spp. using phages. This was achieved by:

1. Establishing a collection of diverse *Klebsiella* isolates (**Chapter 3**)
2. Establishing a phage collection using a range of *Klebsiella* spp. (**Chapter 4**)
3. Characterisation of the phages and appropriate selection of biocontrol phages (**Chapter 4** and **Chapter 6**)
4. Development and implementation of a protocol to produce high-quality, high-titre phage preparations (**Chapter 5**)
5. Generate complete, accurate, and high-quality phage genomes and perform comparative genomic analyses (**Chapter 6**)
6. Generate a transposon mutant library to investigate bacterial genes involved during phage infection (**Chapter 7**)

CHAPTER 2: MATERIALS AND METHODS

2.1 Bacterial strains and growth conditions

2.1.1 General growth conditions

The *Klebsiella* spp. and other bacterial strains used in this work were derived from previous studies (Woodford *et al.*, 2008; Lee *et al.*, 2012; Shin *et al.*, 2012; Chen *et al.*, 2016; Chen *et al.*, 2020) (see **Chapter 3**). *E. coli* BW25113 is the parent strain from the Keio collection (Grenier *et al.*, 2014), and was chosen as the control for the minimum inhibitory concentration assays since it does not contain chromosomally- or plasmid-encoded antimicrobial resistance mechanisms (Baba *et al.*, 2006). For the TraDIS experiments, *K. michiganensis* M7 21 2 #35 wild-type (WT) was selected as the host for which the most phages in the collection were able to infect.

All *Klebsiella* and *E. coli* strains were cultured to stationary phase overnight in brain heart infusion (BHI) agar (Oxoid) at 37°C, unless otherwise specified. Single colony variants (SCVs) of *Klebsiella* spp. were identified on solid media by changes in colony morphology. SCVs were purified by three successive rounds of single colony growth on MacConkey no. 3 agar (Oxoid), and incubated overnight at 37°C.

The *Klebsiella* strain collection was expanded with new isolates cultured from environmental and wastewater samples. 100 μL of each sample was spread on BHI plates and cream-coloured, mucoid colonies were selected, and from there streaked onto MacConkey no. 3 agar (Oxoid). Single, lactose-positive mucoid colonies were selected for three rounds of purification on MacConkey no. 3 agar, and incubated at 37°C. Isolated lactose-positive single colonies were cultured overnight on *Klebsiella* ChromoSelect selective agar (Sigma-Aldrich) to identify putative *Klebsiella* spp.

All bacteria were collected for long term storage by adding purified colonies in 500 μL BHI broth and 500 μL 40% glycerol (final concentration v/v 20% glycerol) and stored at -80°C.

Liquid cultures for phage infections were prepared by inoculating 10 mL of BHI broth with a single colony of bacteria and incubated at 37°C with shaking at 200 r.p.m. for 2-3 h, until at mid-to-late exponential phase ($\approx 10^9$ colony forming units per millilitre, CFU/mL).

2.1.2 Bacterial enumeration

CFU/mL for bacterial strains was calculated following serial dilution of bacterial cultures. Briefly, bacterial cultures were grown to an optical density of 0.4 at 600 nm (OD_{600}) and duplicate or triplicate samples taken and serially diluted ten-fold. 10 μL of serial dilutions were plated onto relevant media. Colonies were counted following incubation at 37°C for 18 hours using the following equation:

$$\mathrm{CFU/mL} = \frac{\mathrm{No.\,of\,colonies}}{\mathrm{Dilution\,factor} \times \mathrm{volume\,(mL)}}$$

The mean for each set of replicates was then calculated.

2.1.3 Generation of phage-resistant mutants

K. michiganensis Oda^R was generated by growing high titres of phage Oda with the *K. michiganensis* M7 21 2 #35 WT bacterial strain and isolating single colonies within the phage plaques. Briefly, 100 µL phage Oda was applied at high titre (10^{10}-10^{11}) to *K. michiganensis* by double agar overlay method (**see section 2.4**) as previously described (Adams, 1959; Kropinski *et al.*, 2009). A single colony within the area of plaquing was purified by three successive rounds of single colony growth on BHI agar.

2.2 Molecular biology

2.2.1 Identification of putative *Klebsiella* spp.

Bacteria preliminarily identified as *Klebsiella* spp. by ChromoSelect agar were sent for 16S rRNA gene sequencing prior to whole-genome sequencing (WGS).

A colony PCR was performed to amplify the 16S rRNA gene. Briefly, a sterile toothpick was used to scrape a small amount of a single colony of purified bacterial culture and resuspended in 200 µL nuclease-free water. Bacterial suspensions were heated to 55°C to lyse the cells.

A 25 µL PCR was performed per sample, consisting of 12.5 µL Invitrogen™ Platinum™ SuperFi™ PCR 2x Master Mix (ThermoFisher) with 1 µL of universal primers U27F (5'-AGAGTTTGATCCTGGCTCAG-3') and U1492R (5'-GGTTACCTTGTTACGACTT-3') each, 2 µL of bacterial suspension template, with the remaining volume made up to 25 µL with nuclease-free water. A Biometra thermocycler was used for the PCR, and conditions were as follows:

> 95°C for 5 minutes
> 30 cycles of   95°C for 30 s
>   62°C for 30 s
>   72°C for 90 s
> 72°C for 10 m
> Hold at 4°C

The PCR products were visualised by agarose gel electrophoresis to check the size of the PCR product. 4 µL of marker Quick-load® 1 kb DNA ladder (NEB) was loaded into the first well. 1 µL 6x gel loading dye (no SDS) (NEB) was added per 5 µL PCR product and 5 µL loaded onto the gel. DNA was resolved by 1% agarose gel electrophoresis in 1X TAE buffer (20 mL of 50X Tris-acetate EDTA buffer (ThermoFisher) made up to 1 L with ultrapure water)

and visualised by Midori Green Xtra (GENEFLOW) staining in a UV cabinet. Visualisation of a band at 1.5 kb was determined as presence of the 16S rRNA gene.

DNA purification of PCR products was performed using QIAquick PCR Purification kit (QIAGEN) and quantified by Qubit 3.0 fluorometer using the broad range dsDNA kit (Invitrogen) according to the manufacturer's instructions. DNA was normalised to 5 ng/μL and genus identification was confirmed by Sanger sequencing (Eurofins) using U27F and U1492R primers.

2.2.2 Bacterial DNA extraction and whole-genome sequencing

DNA for in-house WGS of *Klebsiella* strains was extracted using one of two methods. The first method was with the AllPrep Bacterial DNA/RNA/Protein kit (Qiagen) according to the manufacturer's instructions. The second method was with the Maxwell® RSC Cultured Cells DNA Kit (Promega), where bacterial cultures were grown to mid-to-late exponential phase as described above, and 400 μL of culture added directly to the cartridge according to manufacturer's instructions for Gram-negative cells.

DNA was quantified with a Qubit 3.0 fluorometer (Invitrogen) using the broad range dsDNA kit (Invitrogen) and each sample was normalised to 5 ng/μL. DNA was prepared for sequencing using the Illumina DNA Prep library (formerly Nextera Flex) preparation kit and whole-genome sequenced on the Illumina NextSeq500 platform, generating 2x150 bp paired-end reads by QIB Sequencing Core Services.

Additionally, *K. michiganensis* M7 21 2 #35 WT, *K. pneumoniae* M26 18 2 1, *K. pneumoniae* M26 18 2 #21 KpnN, *K. pneumoniae* M26 18 2 #21 KpnA, and *K. pneumoniae* ST38 01 were prepared for enhanced sequencing by MicrobesNG whole genome sequencing services (www.microbesng.com), which is supported by the Biotechnology and Biological Sciences Research Council (BBSRC). Bacterial samples were prepared according to the sequencing facility's instructions. Briefly, bacterial cultures were grown overnight in BHI broth as previously described, and cells centrifuged at 6,000 x *g* until 0.3-0.6 g of cellular biomass was achieved. Cells were sent to MicrobesNG, and genomes were received from MicrobesNG assembled and annotated.

2.2.3 Plasmid extraction

To create single-gene knockout mutants in *K. michiganensis* M7 21 2 #35 WT, two plasmids were needed from two parent strains: pSIM18 plasmid from its parent strain *S.* Typhimurium SL144, and pKD4 from its parent strain *E. coli* BW25141.

S. Typhimurium SL144 pSIM18 was cultured in 10 mL 0.5% NaCl lysogeny broth (LB) (Oxoid), supplemented with of Hygromycin B Gold (InvivoGen) at a final concentration of 100 μg/mL. pSIM18 contains a temperature-sensitive promoter that required cultures to be grown at 30°C only, with shaking at 200 r.p.m. overnight. *E. coli* BW25141 pKD4 was

cultured in 10 mL BHI supplemented with kanamycin at a final concentration of 50 µg/mL and incubated at 37°C with shaking at 200 r.p.m. overnight.

Plasmids were extracted using the PureYield™ Plasmid Miniprep System (Promega) for larger culture volumes according to the manufacturer's instructions. Plasmid DNA was quantified using Qubit 3.0 fluorometer (Invitrogen) using the high sensitivity dsDNA kit (Invitrogen). Plasmid DNA that was not used immediately was stored at -20°C.

2.2.4 Competent cells and transformations

To make electrocompetent *K. michiganensis* M7 21 2 #35 WT cells for creating knockout mutants, cultures were prepared at room temperature and without ethylenediaminetetraacetic acid (EDTA) supplementation. As no glycerol was used in this protocol, cells needed to be prepared and used the same day and could not be stored for future use.

Stationary phase bacterial cultures were diluted 1:100 and 1.4 mL aliquots were made in 1.5 mL tubes. Tubes were placed into a thermomixer (Eppendorf) at 37°C with 900 rpm for 1 hour 30 minutes (1:30 hours) to 2:00 hours until at mid- to late-exponential phase (OD_{600} = 0.4-0.8).

Cultures were centrifuged at 6,000 x *g* (Eppendorf) for 5 minutes and supernatant discarded. Cells were washed three times by resuspending in 1 mL sterile ultrapure water. For the final step, cells were centrifuged for 10 minutes, as encapsulated cells were not adherent enough otherwise. As much supernatant was removed from each tube as carefully as possible and gently resuspended in 40 µL sterile ultrapure water.

For best transformation efficiency, plasmid concentrations between 1 and 10 ng/µL were used. For pSIM18 at 18 ng/µL, the plasmid DNA was diluted to <10 ng/µL and 1 µL plasmid DNA was added to each 40 µL cell suspension. Cells and DNA suspensions were transferred to 2 mm electrode gap cuvettes (GENEFLOW) and electroporated at 2.5 kV using the MicroPulser Electroporator (BioRad) with the following parameters: voltage at 2500 HV, resistor at 250 Ω, and capacitor at 25 µF. Cells were immediately recovered in 900 µL SOC media (NEB), transferred back into their original 1.5 mL tubes and incubated at 30°C for 1:30 to 2:00 hours.

Following recovery, 100 µL transformants from three tubes were plated onto 0.5% LB agar (Oxoid) supplemented with hygromycin at a final concentration of 100 µg/mL for counting colonies. The remaining 800 µL were centrifuged at 5,000 x *g* for 2 minutes. Most of the supernatant was discarded and pellets resuspended in ≈100 µL of remaining supernatant to reduce the volume for plating. Cells were plated onto 0.5% NaCl LB agar supplemented with hygromycin at a final concentration of 100 µg/mL for counting colonies. Centrifugation and resuspension steps were repeated for all remaining tubes and plated onto 0.5% NaCl

LB agar supplemented with hygromycin at a final concentration of 100 µg/mL. All plates were incubated overnight at 30°C.

Colonies of successful transformants were counted and all colonies were collected in 200-500 µL 0.5% NaCl LB. Glycerol stocks were prepared by adding equal volume of 40% glycerol in sterile ultrapure water to a final concentration of 20% glycerol and stored at -80°C.

K. michiganensis M7 21 2 #35 pSIM18 cells were then ready for creating single gene knockout mutants, which were to be used to validate gene hits identified during the TraDIS experiments.

2.3 Bacterial genomics

2.3.1 Bacterial assembly and annotation

Short-read sequencing data was provided through the QIB Integrated Rapid Infectious Disease Analysis (IRIDA) platform (https://irida.ca/) without pre-processing by QIB Sequencing Core Services. The paired-end reads were QC filtered, trimmed, assembled, and analysed using the ASA³P v1.2.2 (Schwengers *et al.*, 2020) or Bactopia v1.6.4 (Petit and Read, 2020) pipelines and performed by QIB Core Bioinformatics. Both pipelines were used with default settings, and ASA³P was launched using https://github.com/quadram-institute-bioscience/start-asap. Previously sequenced and assembled bacterial genomes were annotated with Prokka v1.14.6 (Seemann, 2014) with default settings on the QIB high performance computer (HPC).

2.3.2 MLST and virulence factor data mining

Preliminary taxonomic assignments for each strain were determined by ribosomal multilocus sequence typing (rMLST) (https://pubmlst.org/species-id) (Jolley *et al.*, 2012). The PubMLST database (https://pubmlst.org/) (Jolley *et al.*, 2018) was used to determine sequence types (ST) for the *K. oxytoca* species complex and *K. aerogenes*, while the Institute Pasteur MLST database (https://bigsdb.pasteur.fr/) was used to determine STs of the *K. pneumoniae* species complex against seven housekeeping genes (*rpoB*, *gapA*, *mdh*, *pgi*, *phoE*, *infB*, *tonB*). Additionally, the chromosomally-encoded bla_{OXY} variants have evolved at a similar rate to MLST housekeeping genes within the *K. oxytoca* species complex and were also included in preliminary taxonomic assignment (Fevre *et al.*, 2005).

The capsular type, O-antigen type, and presence of chromosomal antimicrobial resistance determinants for each strain was predicted using Kleborate v2.0.1 (Lam *et al.*, 2021) and Kaptive (Wick *et al.*, 2018) on the QIB Galaxy platform, and those with a match confidence of good or higher were included.

Virulence genes were predicted using the VFanalyzer platform (http://www.mgc.ac.cn/VFs/), which utilises multiple methods of protein prediction of

virulence factors instead of relying solely on BLAST searches (Liu *et al.*, 2018). All genes present in virulence loci were recorded as present, those missing a minimum of one gene were recorded as incomplete, and those missing all genes were recorded as absent.

2.3.3 Antimicrobial resistance and plasmid data mining

Plasmid and acquired antimicrobial resistance (AMR) genes were predicted with ABRicate v0.9.7 (Seemann, 2016) using the PlasmidFinder (Carattoli *et al.*, 2014) and ResFinder (Bortolaia *et al.*, 2020; Florensa *et al.*, 2022) databases, respectively on the QIB Galaxy platform. Minimum prediction thresholds were set at 90% identity and 80% coverage.

2.3.4 Prophage and anti-phage defence systems data mining

Prophage genome prediction was performed using geNomad v1.6.1 (Camargo *et al.*, 2023) on the QIB HPC. Completeness of geNomad-predicted prophages was performed with CheckV v1.0.1 (Nayfach *et al.*, 2021) on the QIB HPC. Minimum prophage scores were set at 0.7 by default, and only prophages integrated into the genome were included, all other proviral hits were excluded. Anti-phage defence systems were predicted using DefenseFinder v1.1.1 (Tesson *et al.*, 2022) on the QIB HPC. Unique defence system subtypes were included as independent hits, for example, PD-T4-3 and PD-T4-5 were included as two defence systems.

2.3.5 Bacterial comparative genomics

To confirm strain designations, average nucleotide identity (ANI) analysis was performed using FastANI v1.1 (Jain *et al.*, 2018) on the QIB Galaxy platform and visualised with ComplexHeatmap (Gu *et al.*, 2016; Gu, 2022) and heatmap3 (Zhao *et al.*, 2014) in RStudio v2023.12.0+369 using R v4.3.2.

A pangenome was constructed with Roary v3.10.2 (Page *et al.*, 2015) on the QIB Galaxy platform with minimum percentage identity for sequence comparisons performed by protein BLAST (BLASTp) (https://blast.ncbi.nlm.nih.gov/Blast.cgi) set to 80% (default 95%) and threshold of isolates required to define a core gene set at 90% (default 99%) and visualised using Phandango v1.3.0 (https://jameshadfield.github.io/phandango/#/) (Hadfield *et al.*, 2017).

Phylogeny for *Klebsiella* isolates and publicly available *Klebsiella* spp. downloaded from the NCBI database (https://www.ncbi.nlm.nih.gov/) was determined from the DNA sequences of the core genes using IQ-TREE v1.6.12 (Nguyen *et al.*, 2015) on the QIB Galaxy platform. A maximum-likelihood (ML) tree was constructed using ultrafast bootstrapping at 1000 replicates (Hoang *et al.*, 2017). Modelling parameters were determined with the TEST function (Kalyaanamoorthy *et al.*, 2017) and the best-fit substitution model used was GTR+F+I+G4. A core genome multisequence alignment (MSA) of 73 DNA sequences with 1,570,844 nucleotide sites were included in the final analysis. Tree rendering and heatmap

of genomic metadata (ST, capsular type, O-antigen type, AMR, prophage and defence systems) were performed with iTOL v6.1.1 (https://itol.embl.de/) (Letunic and Bork, 2021).

2.4 Isolation and single-plaque purification of phages

Water samples were taken from various rivers and wastewater treatment plants in England and screened for *Klebsiella*-specific phages using an array of *Klebsiella* spp. as hosts, adapted from Van Twest and Kropinski (2009) and Adams (1959). A total of 26 phages were isolated against a range of clinical and environmental *Klebsiella* spp. Briefly, enrichments were performed by inoculating 5 mL BHI broth with premixed 300 µL filtered wastewater and 60 µL exponential growth bacterial culture. Enrichments were incubated overnight at 37°C with shaking at 200 r.p.m. Enrichments were then centrifuged at 4,000 x *g* for 15 minutes and passed through a 0.45 µm filter before spot testing by double agar overlay method. All incubations for the overlay method were performed over 4:00-18:00 hours at 37°C.

To make stocks of single phage strains, three rounds of single-plaque purification were performed. Single plaques were picked from the soft agar layer using sterile toothpicks and suspended in approximately 300 µL BHI broth. Suspensions were vortexed and centrifuged at 13,000 x *g* (Eppendorf) for 5 minutes and supernatant collected. Plaque assay was performed by ten-fold serial dilutions of the supernatant in phage buffer (75 mM NaCl; 10 mM MgSO$_4$; 10 mM Tris, pH 7.5; 0.1 mM CaCl$_2$) and 10 µL of each dilution plated onto double agar overlay.

2.4.1 Phage enumeration

Plaque forming units per millilitre (PFU/mL) for phage enumeration was calculated following serial dilution of phage lysates by plaque assay. Briefly, phage lysates were serially diluted ten-fold and 10 µL of serial dilutions were plated by double agar overlay method, as described above. Plaques were counted following incubation at 37°C for 4:00-18:00 hours using the following equation:

$$\text{PFU/mL} = \frac{\text{No. of plaques}}{\text{Dilution factor} \times \text{volume (mL)}}$$

2.4.2 Multiplicity of inoculation

The multiplicity of inoculation (MOI) refers to the number of infective viral particles relative to the number of host cells. MOI is calculated from the PFU/mL and CFU/mL:

$$\text{MOI} = \frac{\text{PFU/mL}}{\text{CFU/mL}}$$

2.5 Phage amplification

2.5.1 Small-scale solid media phage amplification

Phage amplification was performed as for single-plaque purification in BHI broth. Once supernatant was collected, approximately 100 µL of phage suspension was spread onto double agar overlay plates and incubated as before. Phage stocks were prepared by extraction of phage clearance zones. This was achieved by removal of the soft agar layer, which was resuspended in phage buffer, and centrifuged at 4,000 x g for 15 minutes. Phage supernatant was passed through a 0.22 µm filter into a sterile glass tube and stored at 4°C.

2.5.2 Small-scale liquid phage amplification

For slightly larger volumes, phage amplifications were performed in liquid culture. Briefly, 2.5 mL of stationary phase bacterial culture was used to inoculate 250 mL BHI broth. Cultures were incubated at 37°C with shaking at 200 r.p.m. until at mid exponential phase (OD_{600} 0.1, $\approx 10^7$ CFU/mL). Phage stocks were diluted to an MOI of 1 ($\approx 10^7$ PFU/mL) and 250 µL was added to the bacterial cultures. Cultures were incubated overnight at 37°C with shaking at 160-180 r.p.m.

Stationary phase cultures were centrifuged at 4,000 x g for 15 minutes. The supernatant was passed through a 0.2 µm filter cup attached to a vacuum pump and the bacterial pellets discarded. Phage lysates were stored at 4°C in sterile glass bottles.

2.5.3 Pilot-scale phage amplification by bioreactor

Two types of pilot-scale liquid amplification were performed: pilot-scale amplification by bioreactor and classical pilot-scale amplification.

Pilot-scale phage amplification by bioreactor were performed using the CellMaker Plus Controller (Cellexus) using the regular 4 L CellMaker bioreactor bag. Briefly, 4 L of BHI broth was added to the bioreactor bag with 25 ppm (10 ml) of 1% Antifoam SE-15 (Sigma-Aldrich) at 37°C. Antifoam SE-15 was added as needed, up to a maximum of 100 ppm in 4 L. Air flow was ran at 2-3 litres per minute.

30 mL of stationary phase bacterial culture was used to inoculate the 4 L of BHI broth and incubated for 1:00 hour to obtain $\approx 10^8$ CFU/mL. 8 mL of bacterial culture was removed after 1:00 hour and bacterial enumeration performed as described above, for estimating the MOI. Phage lysate was diluted to obtain an MOI of 0.001 ($\approx 1 \times 10^5$ PFU/mL) and 8 mL added to the 4 L of bacterial culture after 1:00 hour of growth and incubated for 4:00 hours. Phage enumeration for phage lysate dilutions was performed, for estimating the MOI.

Phage lysate was centrifuged at 4,000 x *g* for 15 minutes and supernatant passed through a 0.2 µm filter cup attached to a vacuum pump. Phage lysates were stored at 4°C in glass bottles and phage enumeration was performed as described above.

2.5.4 Classical pilot-scale phage amplification

Volumes of up to a maximum 500 mL of BHI broth was used in two batches of 250 mL to amplify phages. Bacteria were cultured overnight to stationary phase as described above. Cultures were grown to early- to mid-exponential phase (OD_{600} ≈0.2 to 0.25; ≈10^8-10^9 CFU/mL) by adding 2.5 mL of stationary phase culture to 2 x 250 mL BHI broth (1:1000 dilution). Flasks were incubated at 37°C with shaking at 200 r.p.m. for 1:30 to 2:00 hours.

Phage lysates were diluted to obtain an MOI of 0.01-0.1 (≈10^6-10^7 PFU/mL) and 250 µL added to each 250 mL flask. Flasks were incubated at 37°C with shaking at 200 r.p.m. overnight.

Phage lysates were centrifuged at 4,000 x *g* for 15 minutes and supernatant passed through a 0.2 µm filter cup attached to a vacuum pump. Phage lysates were stored at 4°C in glass bottles and phage enumeration was performed as described above.

2.6 Pilot-scale phage concentration and purification

2.6.1 Anion-exchange and fast protein liquid chromatography

Phage lysate was amplified by small-scale liquid amplification as described above. Phage lysate was diluted four-fold (1:4 v/v) in binding buffer (20 mM Tris-HCl, pH 7.5, 0.22 µm filter sterilised) to reduce the ionic load of the lysate (Adriaenssens *et al.*, 2012). A high-affinity CIMmultus® quaternary amine (QA) 8 mL monolithic column (2 µm) (Sartorius) was used for anion exchange and connected to the ÄKTA pure system (GE Healthcare) for fast protein liquid chromatography (FPLC). Additionally, a low-affinity diethylethanolamine (DEAE) column (Sartorius) was used during a second run. The system was sterilised with 1 M NaOH for decontamination followed by sterile ultrapure water. The system was washed with loading buffer until UV absorbance and conductivity had stabilised and phage was run at 8 mL/min.

Phage lysate was loaded onto the system and a stepwise gradient (5% or 10%, 40%, and 100%) of elution buffer (20 mM Tris-HCl, pH7.5, 2 M NaCl, 0.22 µm filter sterilised). Flow through was collected and saved for phage enumeration. Phages were eluted in 6 mL fractions, titrated, and enumerated by double agar overlay method, as described above. Fractions were pooled depending on their UV absorbance and stored at 4°C in glass tubes.

2.6.2 Endotoxin testing

Endotoxin concentration of phage lysates before and after concentration and purification were tested using the Pierce™ Chromogenic Endotoxin Quant Kit (Thermo Scientific)

according to manufacturer's instructions, using the High Standards (0.1-1.0 EU/mL). Phage lysates were diluted by 1×10^{-6} and/or 1×10^{-7} to ensure values were within range, with samples and standards prepared in duplicate in a flat-bottom 96-well plate. Samples and standards were read in a microplate reader at OD_{405} at 37°C. A linear regression standard curve was generated to determine the endotoxin concentration for each sample, where only curves with a coefficient of determination (r^2) of ≥ 0.98 were deemed reliable.

2.6.3 SDS-PAGE of anion-exchange FPLC-purified phages

Phage lysates before and after concentration and purification were assessed for purity relative to one another by denaturation and separation of proteins by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). Protein content of phage lysates before and after concentration and purification were quantified using Qubit 3.0 fluorometer (Invitrogen) using the broad range kit. Briefly, crude lysate was diluted 1:3 before denaturing the proteins. Both the 10% and 40% fractions were loaded at an approximately similar protein concentration, though they could not be normalised due to incompatibility of protein quantification kits and systems.

Protein denaturation, separation, and gel staining was performed according to the manufacturer's instructions for NuPAGE® electrophoresis system (Invitrogen) for reduced samples. Samples were prepared in a total volume of 20 µL with appropriate volume of ultrapure water. Proteins were denatured by adding 5 µL NuPAGE® LDS sample buffer (4X) (Invitrogen) to 13 µL of phage lysate and incubated at 70°C for 10 minutes. 2 µL NuPAGE® Reducing Agent (10X) (Invitrogen) was added to samples just before electrophoresis, and 20 µL total volume was loaded onto a NuPAGE® Novex® 4-12% Bis-Tris Pre-Cast gel, 1.0 mm x 10 well.

Proteins were resolved at 100 V for 1:00 hour in 1X NuPAGE® MES SDS Running Buffer (Invitrogen) for low molecular weight (2-200 kDa) proteins. 500 µL NuPAGE® Antioxidant (Invitrogen) was added to the running buffer in the electrophoresis tank to maintain the proteins in a reduced state.

Gel was stained with InstantBlue® Coomassie Protein Stain (Abcam) according to manufacturer's instructions. Briefly, immersed gel in stain and incubated at room temperature on a rocker at 30 r.p.m. for 1:30 hours.

2.7 Phage infection kinetics

2.7.1 Phage host range

The phage host range was tested, as before, by plaque assay of a dilution series (Kutter, 2005) against a panel of up to 38 *Klebsiella* strains (**see Chapter 4**). Only assays where individual plaques were identified were recorded as positive.

2.7.2 Efficiency of plaquing

Efficiency of plaquing (EOP) was determined was performed by plaque assay using the double agar overlay method as previously described. Efficiency was determined by comparing the tested host strains to the propagation host.

2.7.3 Phage killing kinetics

Phage killing kinetics were determined by plate reader assay. Stationary phase bacterial cultures were adjusted to an overall density of $\approx 10^8$ CFU/mL after growing to mid-exponential phase (OD_{600} of 0.4) and 180 µL added to a CytoOne® flat-bottom 96-well plate, untreated (Starlab). 20 µL of phages serially diluted in phage buffer were added at MOIs ranging from 1000 to 0.00001 at 0 hours in quadruplicate. Plates were sealed with a gas permeable membrane (4TITUDE) and cultures incubated at 37°C with shaking at 500 r.p.m. for 18 hours. OD_{600} was determined at 10-minute intervals in a BioTek LogPhase 600 plate reader and software (Agilent). Data were normalised to media-only controls, which consisted of 180 µL BHI and 20 µL phage buffer.

2.8 Antibiotic synergy assays

2.8.1 Antibiotic susceptibility testing

Antibiotic susceptibility testing was used to determine the minimum inhibitory concentration (MIC) of a compound to completely inhibit bacterial growth. MICs were performed in 96-well plates by microdilution method in cation-adjusted Mueller-Hinton broth (MHB), as previously described (EUCAST, 2000; Wiegand *et al.*, 2008). All antibiotic susceptibility testing was performed with *E. coli* BW25113 as the pan-susceptible control.

Stationary phase *K. michiganensis* M7 21 2 #35 WT and *K. michiganensis* M7 21 2 #35 Oda^R were adjusted to an overall density of $\approx 1 \times 10^5$ CFU/mL. Antibiotics were diluted in MHB by two-fold serial dilutions in duplicate with the following highest concentrations: carbenicillin (CAR), 512 µg/mL; piperacillin (PIP), 32 µg/mL; chloramphenicol (CHL), 32 µg/mL; ciprofloxacin (CIP) 0.125 µg/mL; colistin (CST), 32 µg/mL; cefotaxime (CTX) 0.25 µg/mL; kanamycin (KAN), 32 µg/mL; meropenem (MEM), 0.25 µg/mL; streptomycin (STR), 32 µg/mL; tetracycline (TET), 32 µg/mL; and trimethoprim (TMP), 32 µg/mL. An exception was piperacillin/tazobactam (TZP), where piperacillin concentrations were two-fold and tazobactam concentration was fixed at 4 µg/mL (EUCAST, 2023). Additionally, a tazobactam-only control was performed in conjunction with TZP assays.

Plates were sealed with a gas permeable membrane and incubated statically at 37°C for 18 hours and MICs determined by the lowest concentration of antibiotic with no visible bacterial growth. Additionally, OD_{600} was determined at endpoint in a FLUOstar Omega plate reader

plate reader and software (BMG Labtech). Data were normalised to media-only controls and MICs determined at OD$_{600}$ readings of ≤ 0.06.

2.8.2 Checkerboard assays

Checkerboard assays were used to determine phage-antibiotic synergy, adapted from Nikolic *et al.* (2022). Microdilution of phages and antibiotics were performed in 96-well plates. Antibiotics were diluted two-fold and phages were diluted ten-fold in cation-adjusted MHB across the x- and y-plane, respectively.

Stationary phase *K. michiganensis* M7 21 2 #35 WT was adjusted to an overall density of ≈1x10^5 CFU/mL. Plates were sealed with a gas permeable membrane and incubated statically at 37°C for 18 hours and growth inhibition was determined at endpoint in a FLUOstar Omega plate reader plate reader and software (BMG Labtech). Data were normalised to media-only controls taken before incubation, and growth inhibition determined at OD$_{600}$ readings of ≤ 0.06.

To investigate the interaction effect between antibiotics and phages, the fractional inhibitory concentration (FIC) index was calculated. This represented the sum of the MIC of antibiotic (AC) and the minimum inhibitory MOI (MIM) of phage (BC) in combination divided by the MIC of antibiotic (A) or MIM of phage (B) alone:

$$\text{FIC index} = \text{FIC}_A + \text{FIC}_B = \frac{\text{MIC}_{AC}}{\text{MIC}_A} + \frac{\text{MIM}_{BC}}{\text{MIM}_B}$$

FIC indices were termed: synergistic, ≤ 0.5; additive, > 0.5 to 1; no effect, > 1 to < 2; or antagonistic, ≥ 0.2 (EUCAST, 2000; Nikolic *et al.*, 2022). MIM had to be calculated as for doubling dilutions.

2.9 Phage DNA extraction and whole-genome sequencing

Polyethylene glycol (PEG) precipitation was used to concentrate phage virions prior to DNA extraction. Briefly, DNase I (10 U/µL) (Sigma-Aldrich) and RNase A (10 U/µL) (Sigma-Aldrich) were added at 1 µL for every 1 mL of phage stock and incubated at 37°C for 30 minutes. PEG precipitation was performed with PEG 8000 (10%, w/v) and 1 M NaCl and incubated overnight at 4°C. Precipitate was centrifuged at 17,000 x *g* at 4°C for 10 minutes, supernatant discarded and phage pellet resuspended in 200 µL nuclease-free water. Each 200 µL of phage precipitate was treated with 1 µL proteinase K (50 µg/mL) (Merck), 8 µL 0.5 M EDTA (final concentration 20 mM) and 10 µL 10% SDS (final concentration 0.5%, v/v), and incubated at 55°C for 1:00 hour.

DNA was extracted using the Maxwell® RSC Viral Total Nucleic Acid Purification kit (Promega) according to manufacturer's instructions and eluted in 50 µL nuclease-free water. Phage DNA was quantified with a Qubit 3.0 fluorometer (Invitrogen) using the high

sensitivity dsDNA kit (Invitrogen). DNA was normalised to 5 ng/µL and prepared using an Illumina DNA Prep Library preparation kit and was whole-genome sequenced on the Illumina NextSeq500 platform generating 2x150 bp paired-end reads by QIB Sequencing Core Services. DNA was submitted for ONT sequencing as is and MinION libraries (ONT) were constructed without shearing using the short fragment buffer. Libraries were loaded onto the R9.4.1 flow cell according to manufacturer's instructions by QIB Sequencing Core Services. Additionally, some phage genomes were long-read sequenced using the R10.4.1 flow cell according to manufacturer's instructions.

2.10 Phage genomics

2.10.1 Phage assembly and annotation

Quality control (QC), pre-processing of reads, and *de novo* phage assembly was performed on the QIB Galaxy platform. The quality of short-read data was checked using fastQC v0.11.8 (Andrews, 2010) and pre-processing of reads was performed with fastp v0.19.5 (Chen *et al.*, 2018) based on the fastQC outputs. Briefly, a hard trim was performed of between two and ten bases on both the front and tail to retain at least a per-base quality of 28.

Long-read data were demultiplexed by QIB Core Bioinformatics following sequencing and quality checked with NanoStat v0.1.0 (De Coster *et al.*, 2018). Flye v2.9 (Lin *et al.*, 2016) was used with default settings as the primary assembler as it was deemed the most accurate and reliable (Chen *et al.*, 2021; Wick and Holt, 2021; Wick *et al.*, 2021). One iteration of long-read polishing was performed with Flye assembly by default. Canu v2.2 (Koren *et al.*, 2017) was used as the secondary assembler when Flye was unable to generate a high-quality genome. Error correction and trimming were performed as part of the default settings when assembling with Flye or Canu.

A new HYbrid and Poly-polish Phage Assembly (HYPPA) workflow was developed and validated, which utilised several iterations of long-read and short-read polishing performed on long-read-only assemblies in a specific order (Elek *et al.*, 2023), and is described in detail in **Chapter 6**.

All polishing steps were performed on the QIB Galaxy platform. Briefly, two long-read polishing iterations were performed using Medaka (Wright and Wykes, 2020) with default settings, using the previous polished data as the input for the next round of polishing. Medaka settings were matched for the appropriate ONT chemistry and basecaller. Two short-read polishing iterations were performed using Polypolish (Wick and Holt, 2022) and then POLCA (Zimin and Salzberg, 2020), with default settings. Raw reads were used for each iteration of long-read polishing, and pre-processed short-reads (as described above) for each iteration of short-read polishing.

Prior to development of the HYPPA workflow, short-read-only, traditional hybrid, and long-read-only assemblies were used to resolve phage genomes. Briefly, Shovill v1.0.4 (Seemann, 2018) was used to assemble short-read-only data following pre-processing as described above. Unicycler v0.4.8.0 (Wick *et al.*, 2017) was used for traditional hybrid assemblies, that is, the use of both long- and short-read data in the assembly. Short-read-only polishing of long-read-only assemblies were attempted with Pilon v1.20.1 (Walker *et al.*, 2014).

Following short-read-only assembly of the przondoviruses, contigs were manually checked for DTRs flanking the genome, but none were found. Next, the assemblies were checked for DTRs using PhageTerm (Garneau *et al.*, 2017). However, incorrect phage termini were detected among the przondoviruses and phage T7 reference genome (accession V01146) because PhageTerm is not well suited for Nextera-based sequence libraries. Primer walking was used in an attempt to resolve the direct terminal repeats (DTRs) of przondoviruses. Two or three primers were designed to "walk" each genome end and are given for successful primer walking in phages Oda and Tokugawa only (**Table 2.1**). 2 µL of each primer was added to 15 µL of phage DNA (5 ng/µL) and Sanger sequenced (Eurofins). Sanger sequencing chromatograms were visualised in Artemis 18.2.0 (Carver *et al.*, 2012).

Table 2.1. List of primers used for primer walking

Phage name	Primer name	Sequence (5' → 3')	Genome end	Strand
Oda	CE3_BR_1*	ATATGGTCTCCGGTATTACCTC	Beginning	Reverse
Oda	CE3_BR_2*	TTGAGACCCTCAGTCTAACC	Beginning	Reverse
Oda	CE3_BR_3	CCGTGTTACAGAGGATGGAAG	Beginning	Reverse
Oda	CE3_EF_4*	TGATGACCGACTTGATGCG	End	Forward
Oda	CE3_EF_5	AAAGGCAGAGGGTAGTGATG	End	Forward
Oda	CE3_EF_6*	CTACTCACTTGTAGCGATTAGG	End	Forward
Tokugawa	CE12_BR_3	AATCTCAACCGCCGTTTCC	Beginning	Reverse
Tokugawa	CE12_EF_5	GGAAGGCCACTTTTGGAAAAC	End	Forward

*, primers used for both phage Oda and Tokugawa.

Assemblies generating multiple contigs were checked for contamination using Kraken 2 v2.1.1 (Wood *et al.*, 2019) on the QIB Galaxy platform. Variant calling was performed using iVar v1.0.1 (Grubaugh *et al.*, 2019) on the QIB Galaxy platform. Some long-read-only contigs were double the genome length and identified through alignments performed with progressiveMauve v2.4.0 (Darling *et al.*, 2010). Contigs were then separated and polished as per the HYPPA workflow.

Phage genome annotation was performed using Pharokka v1.2.1 with default settings (Bouras *et al.*, 2022) on the QIB HPC, where coding sequences were predicted with PHANOTATE.

2.10.2 Phage comparative genomics

Where specified, publicly available phage genomes used for comparative genomics were derived from studies listed in **Appendix 10.1**, and downloaded from the NCBI GenBank database.

Verification of assembly quality was performed using a combination of tools. First, read mapping was performed by mapping the raw short- or long-reads to the assembled genome using Bowtie 2 v2.3.4.3 (Langmead and Salzberg, 2012) or BWA-MEM v0.7.17.1 (Li, 2013), respectively on the QIB Galaxy platform. Read mapping was visualised with IGV v2.7.2 (Thorvaldsdottir *et al.*, 2013).

Assemblies in the reverse orientation were reoriented by reverse complementation of the genome in UGENE v38.0 (Okonechnikov *et al.*, 2012). Assemblies were reorientated to begin at the same start point, based on well-curated reference phages. Visualisation and annotation of repeat regions was performed using the Molecular Biology suite on the Benchling platform (https://www.benchling.com/).

The closest relative for each phage was determined as the top hit according to the maximum score identified by nucleotide BLAST (BLASTn) (https://blast.ncbi.nlm.nih.gov/Blast.cgi) using the "optimised for somewhat similar sequences" BLAST parameter. Genes associated with specific phage families, such as DNA-directed RNAP for the family *Autographiviridae*, were identified and used for preliminary taxonomic assignment (Molineux, 2006; Lavigne *et al.*, 2008; Adriaenssens *et al.*, 2020).

Preliminary alignments were performed using progressiveMauve as described above and/or MAFFT auto option (Katoh and Standley, 2013) on the Benchling platform between the closest relative and phages from the same genera. The intergenomic similarity between phages in the collection and a selection of publicly available related phages was calculated using VIRIDIC on the web server (https://rhea.icbm.uni-oldenburg.de/VIRIDIC/) (Moraru *et al.*, 2020).

Phylogenetic analyses were performed using the hallmark RNAP (for przondoviruses) or the TerL amino acid sequence for all other phages and a selection of publicly available phylogenetically related phages downloaded from the NCBI protein database (https://www.ncbi.nlm.nih.gov/). MSA of RNAP or TerL amino acid sequences was performed using the MUSCLE algorithm (Edgar, 2004) in MEGA X v10.0.5 (Kumar *et al.*, 2018) with default settings. An ML tree was generated with 500 bootstraps using the default Jones-Taylor-Thornton model. The RNAP phylogenetic analysis was performed using 35 amino acid sequences, with a total of 684 positions in the final analysis. Tree image rendering was performed with iTOL v6.1.1 (https://itol.embl.de/) (Letunic and Bork, 2021).

Proteomic tree construction of whole genomes for phages was performed with ViPTree v4.0 (https://www.genome.jp/viptree) (Nishimura *et al.*, 2017) against a reference database of Pseudomonadota-infecting phages.

Linear mapping of coding sequences for draft and final phage assemblies was performed using Clinker v0.0.23 (Gilchrist and Chooi, 2021) on the QIB Galaxy Platform.

2.11 TraDIS

2.11.1 *K. michiganensis* M7 21 2 #35 TraDIS mutant library construction

K. michiganensis strain M7 21 2 #35 (wild-type, WT) was previously isolated from the faeces of a preterm infant (Chen *et al.*, 2020). A TraDIS mutant library was constructed in *K. michiganensis* M7 21 2 #35 with the Tn5 transposon as previously described (Langridge *et al.*, 2009; Yasir *et al.*, 2020) with the kanamycin resistance gene (Tn5+Kan^R) as the selective marker. The amplified transposon was provided by Dr Emma Holden. Briefly, the Tn5+Kan^R was amplified using a single primer 5'-CTGTCTCTTATACACATCT-3' that binds to either mosaic end of the transposon, as previously described (Yasir *et al.*, 2020).

K. michiganensis was cultured to stationary phase as described above. Cultures were grown to early- to mid-exponential phase (OD$_{600}$ ≈0.2 to 0.25) by adding 4 mL of stationary phase culture to 4 x 400 mL 2xYT broth. Each flask was supplemented with 5.6 µL 0.5 M EDTA (final concentration 0.7 mM EDTA) and incubated at 37°C with shaking at 180 r.p.m.

The contents of each flask were decanted into 8 x 50 mL aliquots (total = 32). Cells were centrifuged at 4,000 x *g* for 15 minutes and supernatant discarded. Pellets were kept on ice and washed in 12.5 mL ice-cold 10% glycerol (v/v) in sterile ultrapure water, centrifuged as described above, and supernatant discarded. Pellets were washed in 25 mL ice-cold 10% glycerol, centrifuged, and supernatant discarded. Pellets were resuspended in 1 mL ice-cold 10% glycerol and 8 x 50 mL aliquots were pooled into a single 50 mL tube (total = 4). Cells were centrifuged again and supernatant carefully removed. Pellets were resuspended in 1 mL ice-cold 10% glycerol and 500 µL of resuspensions transferred to 8 x 1.5 mL tubes. Cells were centrifuged at 16,000 x *g* for 2 minutes to ensure cells were more adherent, and the supernatant carefully discarded. Pellets were resuspended in 800 µL in ice-cold 10% glycerol and kept on ice. Cells were then added to fresh sterile 1.5 mL tubes in 50 µL aliquots and kept on ice prior to electroporation.

The transposome was prepared by adding 4 µL EZ-Tn5 transposase to 2 µL Tn5+Kan^R DNA (100-150 ng/µL) prepared immediately before use, diluted in 2 µL 100% glycerol. 0.4 µL transposome was added to each 50 µL aliquot of cells. Electroporations were performed as described above. Transformants were immediately recovered in 1 mL BHI broth, pooled in batches of 5 and incubated at 37°C with shaking at 180 r.p.m. for 2 hours 30 minutes.

Pooled transformants were centrifuged at 4,000 x *g* for 5 minutes and 3 mL of supernatant was discarded. Pellets were resuspended in the remaining 2 mL and two tubes of 2 mL were pooled to a total of 4 mL. The 4 mL of resuspended and pooled transformants were plated onto 245x245 mm bioassay plates (Corning) containing BHI agar supplemented with 50 µg/mL kanamycin (Formedium). Three resuspended and pooled transformants were diluted 1:100 and 100 µL plated onto 20 mL BHI agar plate supplemented with 50 µg/mL kanamycin for mutant library enumeration. All plates were incubated at 37°C overnight.

The following day, colonies were counted on the 20 mL plates, the mean was calculated, and counts were extrapolated for the entire library. 7 mL BHI broth was added to the first bioassay plate and all colonies collected and homogenised. The homogenised cell suspension was poured onto the next plate and process repeated for all plates. 2 mL BHI broth was added as needed to account for drying. Once all colonies had been collected and homogenised, 100% glycerol was added for a final concentration of 20% glycerol and stored in 500 µL aliquots at -80°C.

2.11.2 TraDIS mutant library optimisation

Several optimisation steps were performed to determine and/or increase efficiency of the transformations. A trial run of the library construction was performed prior, using small-scale volumes to test a range of EDTA concentrations and impact of restriction enzyme inhibition.

Transformation efficiencies (transformants/µg DNA) were calculated as number of colonies per plate ÷ ng of DNA x 1000 ng/µg. These experiments revealed an optimal final concentration of 0.7 mM of EDTA. Additionally, TypeOne™ Restriction Inhibitor (Cambio) (5 µg/µL) (added at 5 µg per 50 µL cells) had a detrimental effect on transformation efficiency and a decision was made not to include it in the TraDIS mutant library construction.

Cell density may also impact the transformation efficiency, and this was tested mid-way through the construction of the first TraDIS mutant library batch. A batch of competent cells were diluted 1:1 in in ice-cold 10% glycerol prior to electroporation with transposome. Following electroporation, this batch of diluted cells was pooled and recovered as for the dense cells described above. Cell recovery was also adjusted to increase efficiency from 1:30 to 2:30 hours.

Plating, counting, and harvesting of colonies from diluted cells revealed ~35,000 mutants recovered from a single plate, compared to ~45,000 across 14 plates. Construction of the TraDIS mutant library was repeated using diluted cells as described in **section 2.11.1**, generating ~400,000 mutants. The two libraries were pooled into a final library and stored in 20% glycerol in 1 mL aliquots at -80°C.

2.11.3 TraDIS experiments to investigate phage stress responses

To determine genes involved in response to phage infection, the *K. michiganensis* M7 21 2 #35 TraDIS mutant library was grown in the presence of phages representative of various genera: two members of the genus *Slopekvirus*, one of the genus *Przondovirus*, one unclassified myovirus (class *Caudoviricetes*), and one member of the genus *Jiaodavirus*. Each combination was tested in duplicate.

K. michiganensis M7 21 2 #35 TraDIS mutant library was grown to stationary phase in BHI broth supplemented with 50 μg/mL kanamycin. Cultures were adjusted to an overall density of ≈10^8 CFU/mL after growing to mid-exponential phase (OD_{600} of 0.4) in 9 mL BHI broth supplemented with 50 μg/mL kanamycin. Phages were diluted in phage buffer to ≈10^8 PFU/mL in 1 mL and added to 9 mL bacterial culture in duplicate (final MOI of 1) (run 2).

Cultures were incubated at 37°C with shaking at 200 rpm for 1:30 and 3:00 hours. Bacterial cultures were centrifuged at 4,000 x *g* for 10 minutes and most supernatant discarded. Pellets were resuspended in 400 μL remaining supernatant. DNA was immediately extracted using the Maxwell® RSC Cultured Cells DNA Kit (Promega) as described in **section 2.2.2**. DNA was eluted in 100 μL nuclease-free water and quantified by Qubit 3.0 fluorometer using the broad range dsDNA kit (Invitrogen) according to the manufacturer's instructions. DNA was stored at -20°C prior to sequencing library preparation.

2.11.4 TraDIS experiment optimisation

Initially, phage Oda at an MOI of 10 (run 1) completely collapsed the TraDIS mutant library at both 1:30 and 3:00 hour post-inoculation (hpi). Phage time-kill assays were used to adjust time points and MOIs. Briefly, cultures were adjusted to a higher overall density of ≈10^9 CFU/mL after growing to late-exponential phase (OD_{600} of 0.8-1.0) in 9 mL BHI broth supplemented with 50 μg/mL kanamycin. Phage Oda was diluted to a final concentration of ≈10^8 PFU/mL (MOI of 0.1) and ≈10^7 PFU/mL (MOI of 0.01) (run 3).

Time points were adjusted to 15 minutes (0:15), 30 minutes (0:30), and 1 hour (1:00) and bacterial samples removed from the incubator and processed for DNA extraction as described in **section 2.11.3**. All Phage Oda samples at each MOI and time point were tested in duplicate.

2.12 TraDIS sequencing library preparation

2.12.1 MuSeek tagmentation and DNA purification

A single aliquot of the *K. michiganensis* TraDIS mutant library was partially thawed and 50 μL used to inoculate 10 mL BHI broth and grown overnight. Bacterial genomic DNA (gDNA) was extracted using the Maxwell kit as described in **section 2.2.2** and quantified by Qubit

3.0 fluorometer using the broad range dsDNA kit (Invitrogen) according to the manufacturer's instructions. DNA was normalised to 11.1 ng/µL and tagmentation performed using the MuSeek library preparation kit (ThermoFisher), which utilises the MuA transposase to fragment and tag DNA with MuSeek adapters (**Figure 1.6**). Tagmentation was performed as follows: 4.5 µL (50 ng) of each bacterial gDNA sample was added to a PCR tube with 2.5 µL MuSeek Fragmentation Reaction Buffer and 0.5 µL MuSeek Enzyme Mix, mixed and kept on ice. Samples were incubated at 30°C for 8 minutes in a Veriti thermocycler. Immediately after incubation 22.5 µL nuclease-free water was added to make the samples up to 30 µL total volume.

30 µL of each tagmented gDNA sample was transferred to a 96-well PCR plate. Purification and size selection (300-500 bp) of tagmented gDNA was performed with 45 µL (1.5x sample volume) AMPure XP SPRI bead-based Reagent (Beckman Coulter) according to manufacturer's instructions and mixed thoroughly. Samples were incubated at room temperature for 5 minutes and then placed on a magnetic rack for at least 5 minutes, until all the beads had collected and the supernatant was clear. The supernatant was removed and discarded, being careful not to disturb the collected beads. Keeping the PCR plate on the magnetic rack, the beads were washed twice in 200 µL freshly-prepared 80% ethanol. After the final wash, any residual ethanol was removed carefully, and the plate was left to air dry for at least 5 minutes at room temperature. The plate was removed from the magnetic rack and beads were gently resuspended in 22 µL nuclease-free water. The plate was placed back on the magnetic rack until beads had collected and the supernatant was clear. 15 µL of the supernatant containing the purified tagmented gDNA was transferred to fresh 8-strip PCR tubes.

2.12.2 First PCR step and amplicon purification

A Veriti thermocycler was used for all TraDIS library PCR reactions. Two sets of primers were used for the first PCR step: customised i7 primers, specific to the MuSeek adapters, thereby amplifying only the MuSeek fragments; and Tnp001-biotinylated primer (10576.0 µg/µmole) (Invitrogen), specific to the Tn5 fragments for Streptavidin bead capture later. PCR amplification of MuSeek fragments was performed as follows: 20 µL Platinum™ SuperFi II High-fidelity DNA Polymerase 2X Master Mix (Invitrogen), 0.2 µL biotinylated primer, and 4.8 µL 5 mM customised i7 primer was added to 15 µL tagmented gDNA. The PCR conditions were as follows:

      72°C for 3 minutes

      16 cycles of   98°C for 10 s

                      60°C for 60 s

                      72°C for 20 s

      Hold at 4°C

Purification and size selection of PCR product was performed at 1:1 ratio of DNA to AMPure XP beads. Briefly, 40 μL AMPure XP SPRI bead-based Reagent was added to 40 μL PCR product and washed twice with freshly-made 80% ethanol as described above. DNA was eluted in 48 μL nuclease-free water, and after beads were collected on the magnetic rack, 40 μL was transferred to fresh PCR tubes, being careful not to transfer any beads.

2.12.3 Streptavidin purification

Streptavidin bead capture was performed with the Dynabeads™ kilobaseBINDER™ Kit (Invitrogen). Into a fresh PCR plate, 10 μL (100 μg) of thoroughly resuspended beads were transferred and beads were allowed to collect on the magnetic rack for 2-5 minutes. The supernatant was slowly and carefully removed. The plate was removed from the magnetic rack and beads were resuspended in 40 μL Binding Solution very slowly and carefully to avoid foaming. The plate was placed on the magnetic rack again for 2-5 minutes to collect the beads. The supernatant was slowly and carefully removed, and beads washed a second time with 40 μL Binding Solution, again being careful to avoid foaming. 40 μL of the AMPure XP bead-purified PCR product containing the biotinylated Tn5 fragments was added to the resuspended Dynabeads™. Samples were mixed very slowly and carefully to avoid foaming.

The PCR plate was incubated on a rotator set at 50 rpm at room temperature for 4 hours, with intermittent mixing every hour to prevent the beads from settling. After 4 hours, the plate was placed on the magnetic rack and supernatant carefully removed. The Dynabeads™-DNA complex was washed twice in 40 μL Washing Solution and twice in nuclease-free water. For each washing step, the plate was removed from the magnet, the beads resuspended slowly and carefully, and then beads were collected on the magnetic rack before removing and discarding the supernatant. After the final wash in nuclease-free water, Dynabeads™-DNA complex was eluted in 15 μL nuclease-free water.

2.12.4 Second PCR step and amplicon purification

After amplification of all MuSeek fragments only those containing the Tn5 transposon were purified. The next PCR step was used to amplify the Tn5 fragments within the Dynabeads™-DNA complex. This PCR used two sets of primers: the same customised i7 primers that were used for each DNA sample previously and customised i5 primers specific to the Tn5. Thus, only fragments containing MuSeek adapters and Tn5 transposon were amplified. Each DNA sample contained a different combination of i7 and i5 primers, as previously described (Yasir *et al.*, 2020). PCR amplification was performed as follows: 25 μL Platinum™ SuperFi II High-fidelity DNA Polymerase 2X Master Mix (Invitrogen), 5 μL 5 mM customised i5 primer, and 5 μL 5 mM customised i7 primer was added to 15 μL Dynabeads™-DNA complex. The PCR conditions were as follows:

```
        72°C for 3 minutes
    12 cycles of        98°C for 10 s
                        60°C for 60 s
                        72°C for 20 s
    Hold at 4°C
```

The plate was place on the magnetic rack and beads were allowed to collect for 5-10 minutes. Supernatant containing the sample DNA was removed and added to a fresh PCR plate. Purification and size selection of PCR product was performed at 1:1 ratio of DNA to AMPure XP beads (50 µL) as described above. DNA was eluted in 35 µL nuclease-free water and 30 µL of purified amplified DNA was transferred to a fresh PCR plate.

DNA was quantified by Qubit 3.0 fluorometer using the high sensitivity dsDNA kit (Invitrogen) according to the manufacturer's instructions. Samples were pooled and quantified for sequencing using TapeStation (Agilent). DNA was sequenced on the Illumina NextSeq500 platform using the NextSeq 500/550 High Output Kit v2.5 (75 cycles) by Dr Emma Holden and Dr Muhammad Yasir.

2.13 Transmission electron microscopy

Transmission electron microscopy (TEM) was performed by members of the QIB Advanced Microscopy Facility and JIC Bioimaging teams, using preparation equipment and a FEI Talos™ F200C transmission electron microscope (Thermo Scientific) located at the BioImaging Facility, John Innes Centre. Phage suspensions were provided to the facility staff at a minimum concentration of 10^9 PFU/mL in phage buffer or BHI broth. At the facility, 400-mesh carbon-coated copper grids were glow discharged for 20 seconds at 10 mA. Phage particles were adsorbed to grids for 1 minute and negatively stained with 2% uranyl acetate for 1 minute, with excess liquid removed by wicking after each step. Grids were left to air dry before imaging by TEM. Images were deposited on OMERO.web v5.19.0 and measured using Fiji v1.53c (https://github.com/fiji/fiji).

2.14 Data analysis

All data unless otherwise specified are represented as the mean of at least three technical repeats, and ± standard deviation (SD) error was generated from at least two biological repeats ($n = 2$).

Statistical analyses for EOP were performed by one-way ANOVA corrected for multiple comparison (Tukey method) with GraphPad Prism v10.1.2 as described in figure legend. Statistical significance was defined as $p \leq 0.05$, pairwise comparisons that was not significant were not shown.

Statistical analyses for transformation efficiencies were performed by two-way ANOVA (full model) corrected for multiple comparison (Tukey method) with GraphPad Prism v10.1.2 as described in figure legend. Statistical significance was defined as $p \leq 0.05$, pairwise comparisons that was not significant were not shown.

Initial analysis of the *K. michiganensis* TraDIS mutant library was performed by Dr Muhammad Yasir. Subsequent analyses were performed by myself with the assistance of Dr Emma Holden.

TraDIS sequencing reads were mapped back to the reference *K. michiganensis* M7 21 2 #35 WT genome and analysed with BioTraDIS (Barquist *et al.*, 2016). Transposon insertions were identified by the number of reads that mapped to a particular site against the reference genome. Plot files were visually examined in Artemis 18.2.0 (https://sanger-pathogens.github.io/Artemis/Artemis/) (Carver *et al.*, 2012) and log fold change calculated with BioTraDIS to determine the difference between the control and phage-treated conditions. From these data, genes important for phage infection and potential single-gene knockout candidates were identified and plotted using GraphPad Prism v10.1.2.

Plot files for *K. michiganensis* M7 21 2 #35 mutant library controls were converted to generate the number and frequency of transposon insertions for both the forward and reverse strands and mapped back to the *K. michiganensis* M7 21 2 #35 WT reference genome. The insertional frequency of insertional mutants across the genome were visualised with DNAPlotter v18.2.0 (https://sanger-pathogens.github.io/Artemis/DNAPlotter/) (Carver *et al.*, 2008).

Figures generated with BioRender (https://www.biorender.com/) are given in figure legends.

CHAPTER 3: *KLEBSIELLA* STRAIN COLLECTION AND GENOMICS

3.1 Introduction

The genus *Klebsiella* has a complex population structure comprised of 21 species where the majority fall into one of two species complex (Podschun and Ullman, 1998; Brisse and Verhoef, 2001; Holt *et al.*, 2015). The *K. oxytoca* species complex is comprised of *K. oxytoca* (*sensu stricto*), *K. grimontii*, *K. michiganensis*, *K. spallanzanii*, *K. pasteurii*, and *K. huaxiensis*. The *K. pneumoniae* species complex is comprised of *K. pneumoniae* (*sensu stricto*) of which there are three subspecies, *K. quasipneumoniae* of which there are two subspecies, *K. variicola* of which there are two subspecies, *K. quasivariicola*, and *K. africana*. The remaining six species do not fall into a specific complex.

Klebsiella spp. are unique to most other genera within the order Enterobacterales in that their pathogenicity largely relies on defence mechanisms as virulence factors rather than active mechanisms such as toxin production and motility (except in the case of *K. aerogenes*) (Paczosa and Mecsas, 2016). The most well-characterised virulence factor among *Klebsiella* spp. is the capsule. Within the *K. pneumoniae sensu stricto*, there are more than 79 different capsular serotypes, and many more capsular locus (KL) types that have been genomically characterised (Paczosa and Mecsas, 2016; Lam *et al.*, 2022). The capsule of other species within the genus *Klebsiella* is less well-characterised but adds to the genomic and population complexity of the genus.

The capsule is the first bacterial component that *Klebsiella*-targeting phages interact with, whether that be through utilisation as a primary receptor for reversible adsorption, or as a barrier to overcome through degradation by phage-encoded depolymerases (Cai *et al.*, 2019; Pan *et al.*, 2019). Some phages may encode multiple depolymerases within their tail fibre and/or spike proteins, which can increase their host range across multiple capsular types (Pan *et al.*, 2017; Cai *et al.*, 2019; Pan *et al.*, 2019). Other important virulence factors that may be important during phage infection include many OM-bound proteins, including porins and proteins necessary for nutrient uptake and the O-antigen (Molineux, 2006; Maffei *et al.*, 2021). Indeed, many OM-bound components act as the primary and/or secondary receptor important during phage infection.

Klebsiella spp. are important opportunistic pathogens, being responsible for a range of serious infections including UTIs, pneumonia, and wound infections that can lead to sepsis (Podschun and Ullman, 1998; Holt *et al.*, 2015). Their ability to rapidly disseminate within hospital settings coupled with their ability to form biofilms and extensive AMR, has resulted in significantly recalcitrant infections, for which new therapeutics are desperately needed (Podschun and Ullman, 1998; Holt *et al.*, 2015; Navon-Venezia *et al.*, 2017; Theuretzbacher, 2017; Bengoechea and Sa Pessoa, 2019). Amongst these, there are a number of global problem clones with extensive AMR profiles that may or may not be

associated with additional virulence factors, such as plasmid-mediated siderophores (Paczosa and Mecsas, 2016; Wyres *et al.*, 2019; Wyres *et al.*, 2020).

It is important to understand the genomic background of the bacteria that are used to characterise phages, particularly with regards to host range and phage infection dynamics. Moreover, these data are integral for rational design of phage preparations for use as therapeutics, particularly regarding globally disseminated MDR clones. Additionally, bacteria can encode multiple anti-phage defence mechanisms and their genomes may contain multiple prophages, both of which can affect infection by other virulent phages.

The data presented in this chapter will have the results and discussion combined.

3.1.1 Aims

1. Determine the pangenome and phylogeny of *Klebsiella* spp. in the collection to aid taxonomic assignment
2. Genomically characterise the *Klebsiella* spp. by identifying ST, virulence factor, and AMR profiles

3.2 Genomic analyses of *Klebsiella* strains

To investigate the biocontrol potential of bacteriophages, first a collection of a diverse set of strains belonging to the genus *Klebsiella* was made. Strains were acquired for either the physical collection, or for the genomic collection with the intention of adding non-clonal strains to the physical collection. Some isolates were added to the physical collection and used in phage phenotypic studies for which sequencing data was not available. The strains used in the various analyses are summarised below (**Table 3.1**):

- Total isolates in the physical collection: 52
    1. Total isolates in the physical collection that were genomically characterised: 43
    2. Total isolates in the physical collection that were not genomically characterised: 9
- Total isolates genomically characterised, but not in the physical collection: 19
- Total reference strains: 21

3.2.1 Strain identification, pangenome analyses, and phylogeny for taxonomic assignment

Preliminary taxonomic assignment of all *Klebsiella* genomes was performed with rMLST, except for *K. oxytoca* M59 22 8 KoN and *K. oxytoca* M59 22 8 KoA, for which no sequencing data was found and were re-sequenced.

Table 3.1. Metadata for bacterial strain collection.

Bacterial species	Strain code or description	Source	Physical collection	Genomic collection	Accession	Reference
Clinical isolates						
K. aerogenes	CL1	Child	Yes	No	-	Clinical strain, GOSH
K. aerogenes	CL2	Child	Yes	No	-	Clinical strain, GOSH
K. aerogenes	CL3	Child	Yes	No	-	Clinical strain, GOSH
K. grimontii	P038I	Infant faeces	Yes	Yes	QFVM00000000	(Chen *et al.*, 2020)
K. grimontii	P043G P	Infant faeces	Yes	Yes	QFVL00000000	(Chen *et al.*, 2020)
K. michiganensis	M7 21 2 #35 WT	Infant faeces	Yes	Yes	-	Hall Lab collection
K. michiganensis	M7 21 2 #35 Oda^R	Derived from WT	Yes	No	-	This study
K. michiganensis	P049A W	Infant faeces	Yes	Yes	QFUG00000000	(Chen *et al.*, 2020)
K. michiganensis	P095L Y	Infant faeces	Yes	Yes	QFTT00000000	(Chen *et al.*, 2020)
*K. oxytoca**	M59 22 8 KoN	Infant faeces	Yes	Yes	-	Hall Lab collection
*K. oxytoca**	M59 22 8 KoA	Derived from KoN	Yes	Yes	-	This study
K. oxytoca	CL4	Child	Yes	No	-	Clinical strain, GOSH
K. oxytoca	CL5	Child	Yes	No	-	Clinical strain, GOSH
K. pneumoniae	CL6	Child	Yes	No	-	Clinical strain, GOSH
K. pneumoniae	CL7	Child	Yes	No	-	Clinical strain, GOSH
K. pneumoniae	M26 18 1	Infant faeces	Yes	Yes	-	Hall Lab collection
K. pneumoniae	M26 18 2 #21 KpnN	Derived from M26 18 1	Yes	Yes	-	Hall Lab collection
K. pneumoniae	M26 18 2 #21 KpnA	Derived from KpnN	Yes	Yes	-	This study
K. pneumoniae	NCTC13438	NCTC	Yes	Yes	UGKS01000005	(Woodford *et al.*, 2008)
K. pneumoniae	P008E 1	Infant faeces	Yes	Yes	QFUX00000000	(Chen *et al.*, 2020)
K. quasipneumoniae	P057K W	Infant faeces	Yes	Yes	QFVB00000000	(Chen *et al.*, 2020)

K. variicola	CL8	Child	Yes	No	-	Clinical strain, GOSH
Food isolates						
K. michiganensis	0371-2842-80A	Food	No	Yes	-	Kingsley Lab collection
K. pasteurii	H-27-MC418-422col86	Food	No	Yes	-	Kingsley Lab collection
K. pneumoniae	0287-A1	Food	No	Yes	-	Kingsley Lab collection
K. pneumoniae	0287-B4	Food	No	Yes[A]	-	Kingsley Lab collection
K. pneumoniae	0287-C2	Food	No	Yes	-	Kingsley Lab collection
K. pneumoniae	0287-F3	Food	No	Yes	-	Kingsley Lab collection
K. pneumoniae	0287-G3	Food	No	Yes	-	Kingsley Lab collection
K. pneumoniae	0287-H1	Food	No	Yes	-	Kingsley Lab collection
K. pneumoniae	0293-627-71B	Food	No	Yes	-	Kingsley Lab collection
K. pneumoniae	0310-357-17A	Food	Yes	Yes[A]	-	Kingsley Lab collection
K. pneumoniae	0310-357-20A	Food	Yes	Yes[A]	-	Kingsley Lab collection
K. pneumoniae	0311-342-6	Pork meat product	Yes	Yes	SAMN33593349[B]	(Elek *et al.*, 2023)
K. pneumoniae	0311-352-13	Food	No	Yes	-	Kingsley Lab collection
K. pneumoniae	0366-2797-39	Food	Yes	Yes[A]	-	Kingsley Lab collection
K. pneumoniae	0419-377-36	Food	No	Yes	-	Kingsley Lab collection
K. pneumoniae	2023-2027-67	Food	Yes	Yes	-	Kingsley Lab collection
K. pneumoniae	H-26-MC2128-2132-75Bm	Food	No	Yes	-	Kingsley Lab collection
K. pneumoniae	H-26-MC2393-2397-33	Pork meat product	Yes	Yes	SAMN33593350[B]	(Elek *et al.*, 2023)
K. pneumoniae	H-26-MC2393-2397-34	Food	Yes	Yes[A]	-	Kingsley Lab collection
K. pneumoniae	H-26-MC2393-2397-35	Food	Yes	Yes[A]	-	Kingsley Lab collection
K. pneumoniae	H-27-MC778-782col36	Food	Yes	Yes	-	Kingsley Lab collection
K. pneumoniae	H-27-MC778-782col37	Food	Yes	Yes[A]	-	Kingsley Lab collection
K. pneumoniae	H-27-MC778-782col38B	Chicken meat product	Yes	Yes	SAMN33593351[B]	(Elek *et al.*, 2023)

Species	ID	Source			Accession	Reference
K. pneumoniae	H-27-MC838-842col2mix	Food	No	Yes	-	Kingsley Lab collection
K. pneumoniae	MC133-137-25	Chicken meat product	Yes	Yes	SAMN33593347[B]	(Elek *et al.*, 2023)
K. pneumoniae	MC148-152-26	Chicken meat product	Yes	Yes	SAMN33593348[B]	(Elek *et al.*, 2023)
K. pneumoniae	MC148-152-27	Food	Yes	Yes[A]	-	Kingsley Lab collection
K. pneumoniae	MC158-162-39	Food	Yes	Yes	-	Kingsley Lab collection
K. pneumoniae	MC223-227-6	Food	Yes	Yes	-	Kingsley Lab collection
K. pneumoniae	MC313-317-22	Food	Yes	Yes	-	Kingsley Lab collection
K. pneumoniae	MC343-352-15	Food	No	Yes	-	Kingsley Lab collection
K. pneumoniae	MC1638-1642-73m	Food	Yes	Yes[A]	-	Kingsley Lab collection
K. pneumoniae	MC1678-1682-23	Food	No	Yes	-	Kingsley Lab collection
K. pneumoniae	MC1968-1972-12mix	Food	No	Yes[A]	-	Kingsley Lab collection
K. variicola	MC223-227-4	Food	No	Yes	-	Kingsley Lab collection
K. variicola	MC283-287-54	Food	No	Yes	-	Kingsley Lab collection
K. variicola	MC1688-1692-38	Food	No	Yes	-	Kingsley Lab collection
Misc. isolates						
K. pneumoniae	ST38 01	Unknown	Yes	Yes	-	Webber Lab collection
K. pneumoniae	ST38 02	Unknown	Yes	Yes	-	Webber Lab collection
Type isolates						
K. aerogenes	DSM 30053	DSMZ	Yes	Yes	CP002824	(Shin *et al.*, 2012)
K. pneumoniae	DSM 30104	DSMZ	Yes	Yes	AJJI01000001	(Lee *et al.*, 2012)
K. variicola	DSM 15968	DSMZ	Yes	Yes	CP010523	(Chen *et al.*, 2016)
Wastewater isolates						
K. huaxiensis	0220.RS02	Raw sewage	Yes	Yes	-	This study
K. michiganensis	0220.ML02	Mixed liquor	Yes	Yes	-	This study
K. michiganensis	0220.ML04	Mixed liquor	Yes	Yes	-	This study

K. pneumoniae	0220.ML03	Mixed liquor	Yes	Yes	-	This study
K. pneumoniae	0220.RS01	Raw Sewage	Yes	Yes	-	This study
K. ornithinolytica	0220.AF01	Aeration feed	Yes	Yes	-	This study
K. ornithinolytica	0220.ML01	Mixed liquor	Yes	Yes	-	This study
Reference strains						
K. aerogenes	-	NCBI	No	Yes	GCF_900075225	(Wyres _et al._, 2020)
K. africana	-	NCBI	No	Yes	FF1003	(McDougall _et al._, 2021)
K. electrica	-	NCBI	No	Yes	GCF_002806725	(Wyres _et al._, 2020)
K. grimontii	-	NCBI	No	Yes	GCF_004104525	-
K. huaxiensis	-	NCBI	No	Yes	GCF_902158605	-
K. indica	-	NCBI	No	Yes	GCA_005860775	-
K. michiganensis	-	NCBI	No	Yes	GCF_000240325	(Wyres _et al._, 2020)
K. ornithinolytica	-	NCBI	No	Yes	GCA_900083685	(Wyres _et al._, 2020)
K. oxytoca	-	NCBI	No	Yes	GCF_900636985	-
K. pasteurii	-	NCBI	No	Yes	GCF_000247915	(Wyres _et al._, 2020)
K. planticola	-	NCBI	No	Yes	GCA_900083755	(Wyres _et al._, 2020)
K. pneumoniae s. _oz_	-	NCBI	No	Yes	GCF_003010735	-
K. pneumoniae s. _pn_	-	NCBI	No	Yes	GCF_000775375	-
K. pneumoniae s. _rh_	-	NCBI	No	Yes	GCF_000163455	-
K. quasipneumoniae s. _qp_	-	NCBI	No	Yes	GCF_000492415	(Wyres _et al._, 2020)
K. quasipneumoniae s. _si_	-	NCBI	No	Yes	GCF_000613225	-
K. quasivariicola	-	NCBI	No	Yes	GCF_000523395	(Wyres _et al._, 2020)
K. spallanzanii	-	NCBI	No	Yes	GCF_902158555_SB6411	-
K. terrigena	-	NCBI	No	Yes	GCF_902109485	-
K. variicola s. _tr_	-	NCBI	No	Yes	GCF_001548315	(Wyres _et al._, 2020)

K. variicola s. *va*	-	NCBI	No	Yes	CP000964	-

*, later re-assigned to the species *K. grimontii*. Cyan, 43 isolates that were in both the physical and genomic collection; yellow, 19 additional food strains in the genomic collection but not in the physical collection, giving a total of 62 isolates genomically characterised. Nine isolates were in the physical collection but not genomically characterised, giving a total of 52 physical isolates. Twenty-one isolates were reference strains. [A], ten isolates could not be included in the phylogenomic tree but were in all other analyses (**Appendix 10.2**), giving a total of 73 isolates including reference strains included in the phylogeny. [B], specific food source given and overall BioProject accession PRJNA941224, no metadata was available for the other food samples; -, no data; s. *oz*, subspecies *ozaenae*; s. *pn*, subspecies *pneumoniae*; s. *rh*, subspecies *rhinoscleromatis*; s. *qp*, subspecies *quasipneumoniae*; s. *si*, subspecies *similipneumoniae*; s. *tr*, subspecies *tropica*; s. *va*, subspecies *variicola*. GOSH, Great Ormond Street Hospital

To investigate the taxonomic assignment of *Klebsiella* isolates, the average nucleotide identity (ANI) was calculated between all 62 sequenced draft genomes and 21 reference genomes (*n* = 83) (**Figure 3.2**). Reference genomes were selected based on other robust studies. Species ANI clustering thresholds (Konstantinidis and Tiedje, 2005; Arahal, 2014) were defined as:

- Within the same species: ≥ 95%
- Clonal species: ≥ 99.9%
- Identical species: 100%

Data for the genus *Salmonella* suggest that a ≥ 98% ANI is sufficient to distinguish subspecies (Pearce et al., 2021), but is less clear for the genus *Klebsiella* and not defined in this study. Clonal species threshold was arbitrarily selected as ≥ 99.9% ANI.

All species were defined clearly and clustered within their species complex, i.e. preliminary rMLST had assigned each isolate to the correct species, which was confirmed by the ANI analyses. However, two strains originally thought to be *K. oxytoca* based on previous metadata were revealed to be *K. grimontii* based on the ANI analysis (> 99% ANI with *K. grimontii* reference strain), and not the species *K. oxytoca* (91% ANI with *K. oxytoca* reference strain). No preliminary rMLST was performed for these two isolates, which are hereafter referred to as *K. grimontii* unless otherwise stated. *K. grimontii* M59 22 8 KoA (hereafter *K. grimontii* KoA) was derived from *K. grimontii* M59 22 8 KoN (hereafter *K. grimontii* KoN) as a spontaneous single colony variant (SCV), and ANI analyses confirmed they were clonal but not identical (> 99.9% ANI).

ANI was used to explore the clonality of isolates of the same species which were isolated from the same wastewater treatment plant: two *K. michiganensis* isolates, two *K. ornithinolytica* isolates, and two *K. pneumoniae* isolates. It was important to establish whether the isolates were clonal, to better assess and interpret phage host range analyses. The two *K. michiganensis* isolates (0220.ML02 and 0220.ML04) were both assigned to the same species, but were not clonal (> 98%, < 99.9% ANI). Similarly, the two *K. ornithinolytica* (0220.AF01 and 0220.ML01) and two *K. pneumoniae* (0220.ML03 and 0220.RS01) isolates were also members of the same species as one another but not clonal (> 99%, < 99.9% ANI).

Most *K. pneumoniae* strains exhibited pairwise nucleotide identities in the range of 98% to > 99.99% (> 98% ANI) when compared to one another. ANI revealed several subpopulations of clonal *K. pneumoniae* where nucleotide similarity was > 99.9%, which was particularly common among many food isolates. Both *K. pneumoniae* M26 18 2 #21 KpnN and KpnA (hereafter *K. pneumoniae* KpnN and *K. pneumoniae* KpnA, respectively) were derived from *K. pneumoniae* M26 18 1, with *K. pneumoniae* KpnA being a spontaneous SCV of *K. pneumoniae* KpnN, so it was unsurprising that their ANI values were very high (> 99.99%).

The two *K. pneumoniae* ST38 variants also showed high nucleotide similarity and were clonal (> 99.98% ANI).
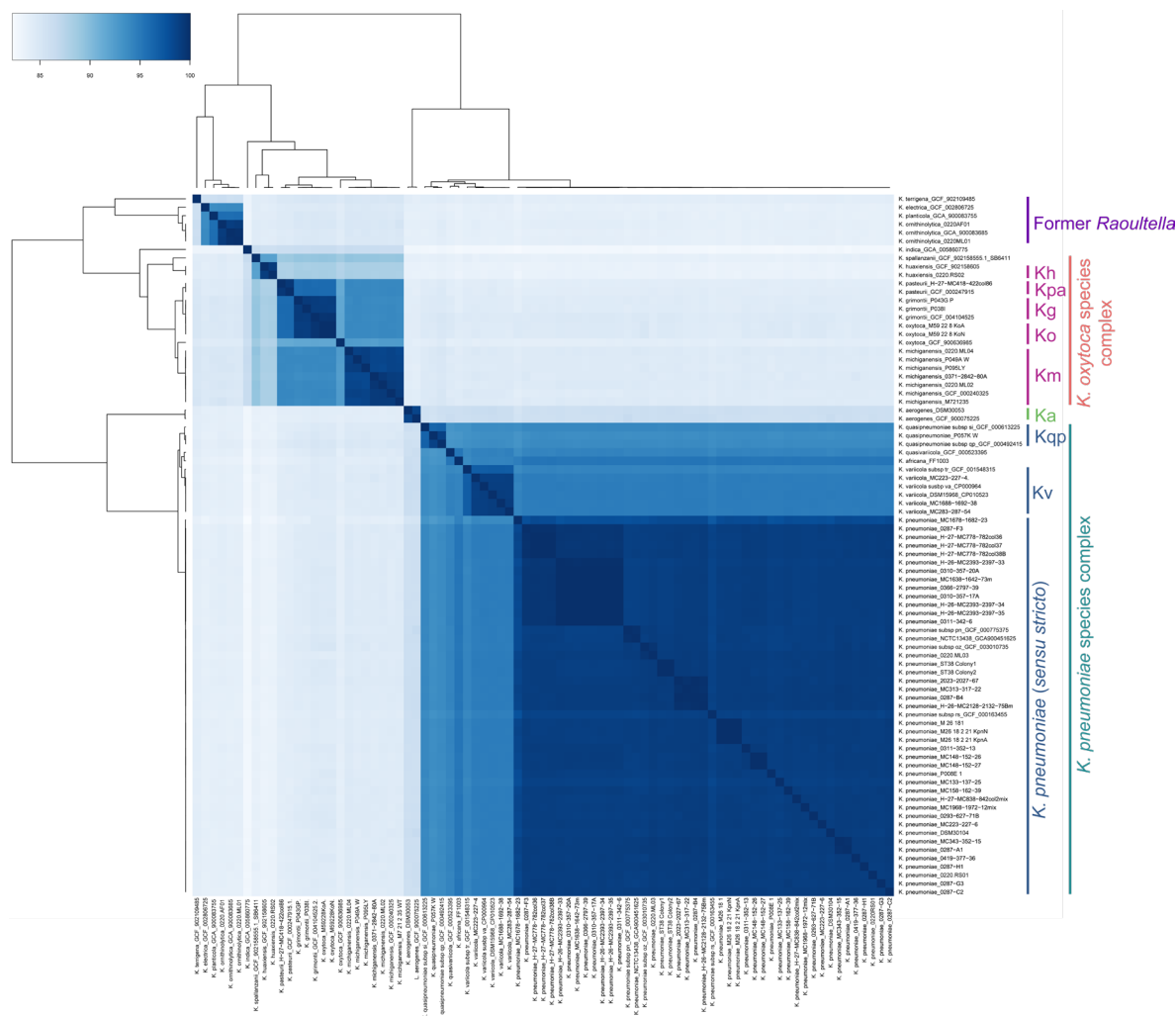
Figure 3.1. Heatmap and dendrogram of average nucleotide identity (ANI) among 83 *Klebsiella* isolates. ANI was performed for *Klebsiella* genomes across almost all species. Dendrogram clustering was based on pairwise nucleotide identities and scale bar represents % ANI, with darker colour denoting higher identity scores. Kh, *K. huaxiensis*; Kpa, *K. pasteurii*; Kg, *K. grimontii*; Ko, *K. oxytoca*; Km, *K. michiganensis*; Ka, *K. aerogenes*; Kqp, *K. quasipneumoniae*; Kv, *K. variicola*; other species not labelled on the plot include *K. indica*, *K. spallanzanii*, K. *quasivariicola*, and *K. africana*.

To explore the plasticity of the *Klebsiella* genomes, a pangenome analysis was conducted with the 83 genomes as selected previously. A total of 31,758 genes were analysed, of which 2,956 were considered core genes (present in > 90-100%) (**Figure 3.2**). Sixty three soft core genes (present in ≥ 89% to ≤ 90%) were identified, along with 3,956 shell genes (present in > 15 to < 89%), and 24,783 cloud genes (present in < 15%). The core genome represented only 9.31% of the pangenome, whereas the shell and cloud genes, together the accessory genome, represented 12.46% and 78.04%, respectively. There was a clear clustering of non-*K. pneumoniae* genomes that showed similar patterns in their gene presence/absence composition, with two clades of *K. pneumoniae* species complex distributed at either end.
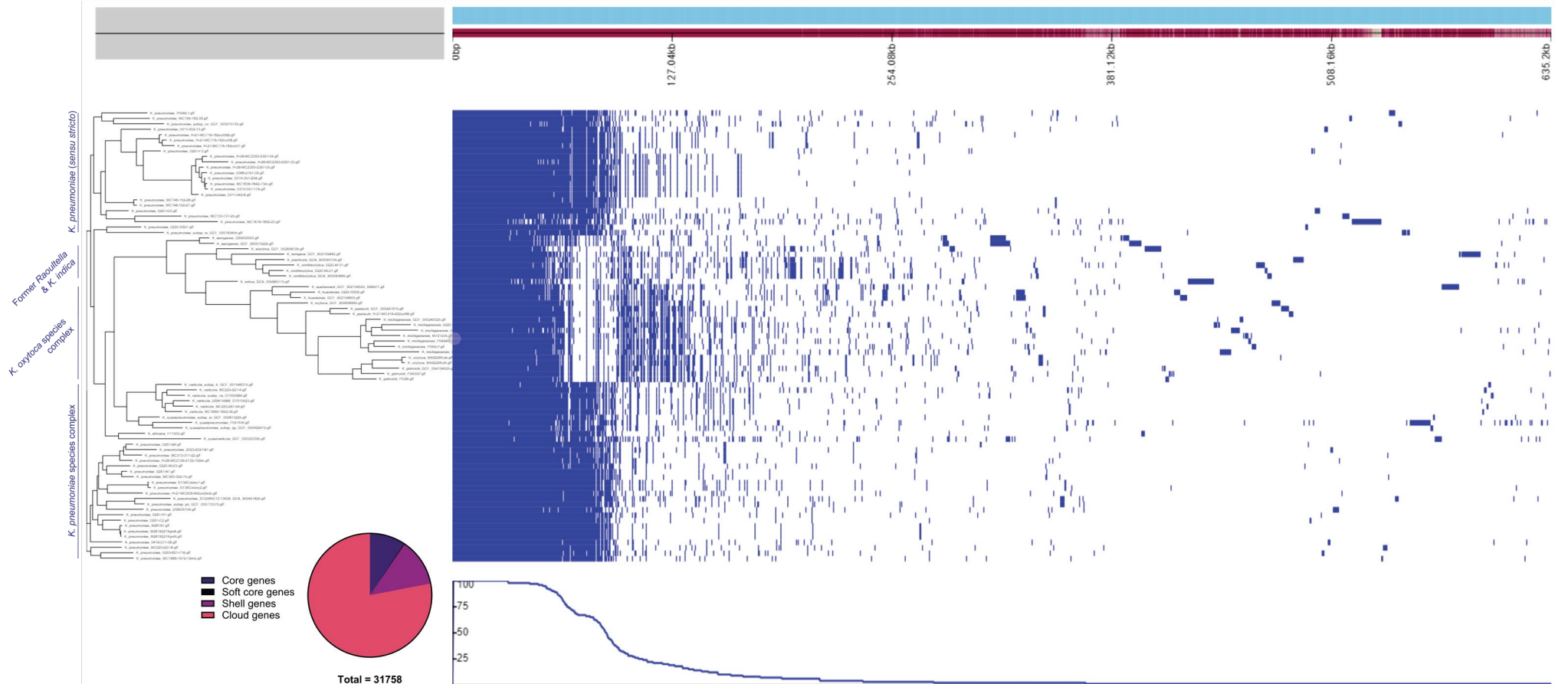
Figure 3.2. Neighbour-joining phylogeny of the presence and absence of genes and distribution of genes across the pangenome of *Klebsiella* strains. Genomes are clustered according to presence (blue) and absence (white) of core and accessory genes, together with pangenome statistics (bottom left). Pie size corresponds to the number of genes in each category with the total number of genes analysed given below. Pangenome analyses (tree and gene presence/absence matrix) were performed with Roary and visualised with Phandango.

The pangenome highlighted clade-specific differences, particularly between different species complexes. Considerably fewer accessory genes were shared among the group *K. electrica*, *K. ornithinolytica*, *K. planticola*, and *K. terrigena* (referred to as "plant-associated" along with *K. indica* and *K. variicola*), than with other *Klebsiella* spp. Additionally, the *K. oxytoca* species complex did not share as many accessory genes with *K. pneumoniae* species complex, or any other *Klebsiella* spp., but did show similarities in accessory genome composition among themselves.

Two distinct clades emerged from the analyses based on the composition of shell and cloud genes shared among them. Non-*K. pneumoniae* species within the *K. pneumoniae* species complex (*K. africana*, *K. quasipneumoniae*, *K. quasivariicola*, and *K. variicola*) showed minor differences in their gene presence/absence profiles to the *K. pneumoniae sensu stricto*. That is, there were some genes present in the latter that were not present in the former. Interestingly, two clades of *K. pneumoniae sensu stricto* emerged, which also showed minor differences in their gene/presence absence composition, though the differences were not consistent among all species. Almost all of the first clade were foodborne *K. pneumoniae*, many of which were clonal subpopulations identified in the ANI analyses.

In *K. pneumoniae*, the majority of the genome is comprised of accessory genes (~3,500 genes, 64%), highlighting the highly adaptable nature of this bacterium (Rocha *et al.*, 2022). Like all *Klebsiella* spp., *K. pneumoniae* is environmentally ubiquitous, but is the most commonly isolated species of the genus within clinical settings (Podschun and Ullman, 1998; Holt *et al.*, 2015; Paczosa and Mecsas, 2016; Merla *et al.*, 2019). Moreover, the ability to occupy multiple niches such as the environment, food, and clinical settings would require high genomic plasticity, which is possibly why *K. pneumoniae sensu stricto* was overrepresented in food samples. Additionally, the differences observed between the gene presence/absence profiles between clusters, such as members of the *K. oxytoca* species complex and *K. pneumoniae* complex species may reflect the different and often overlapping biochemical characteristics, for which the genes they encode may vary greatly (Lin *et al.*, 2015; Fontana *et al.*, 2019).

To explore the phylogeny of *Klebsiella* spp. and check correct species assignment, a phylogenomic tree was constructed from the core genome alignment (**Figure 3.3**), which was generated with Roary and performed separately to the neighbour-joining phylogeny for the gene presence/absence analysis. However, ten genomes could not be included in the phylogenomic analysis, but data mining was performed for which metadata is available (**Appendix 10.2**).

Phylogeny showed four distinct clades: the former *Raoultella* spp.; *K. oxytoca* species complex; *K. aerogenes*; and *K. pneumoniae* species complex. Genomes clustered according to their previously assigned species, as expected. Here, these species were

nested within the genus *Klebsiella*, confirming the genus *Raoultella* as obsolete. In the rest of this work, only the *Klebsiella* terminology will be used.

Genomes within the *K. oxytoca* species complex and *K. pneumoniae* species complex clustered into distinct clades, with each species showing subclades. As expected, the *K. grimontii* M59 22 8 variants, which were previously thought to be *K. oxytoca* clustered with *K. grimontii*. In concordance with the ANI data, less diversity was seen between *K. pneumoniae* genomes, particularly those within the foodborne subpopulations.

3.2.2 Identification of global problem clones

MLST was performed to identify global problem clones within the isolate collection (**Figure 3.3**, **Appendix 10.2**), such as those from *K. pneumoniae* clonal group (CG) 258. MLST identified five different STs within the *K. oxytoca* species complex, with a further five being non-typeable, including the two lesser-known species *K. huaxiensis* and *K. pasteurii*. The isolate of *K. aerogenes* belonged to ST134 (CG1), which is one of the most predominant STs of the species (Manandhar *et al.*, 2022). *K. pneumoniae* isolate NCTC13438 belonged to the globally disseminated CG258 (ST258), and is a well-known global problem clone. The wastewater-derived isolate *K. pneumoniae* 0220.ML03, belonged to the global problem lineage ST37 (CG37) which can be either MDR or non-MDR (Guo *et al.*, 2016; Wyres *et al.*, 2020). However, *K. pneumoniae* 0220.RS01, whilst sharing > 99% ANI with 0220.ML03, belonged to ST1618. Additionally, both *K. pneumoniae* 0220.ML03 and 0220.RS01 were predicted to have different capsular types, KL38 and KL30, respectively. No ST could be determined for either *K. ornithinolytica* isolates and the *K. michiganensis* isolate, 0220.ML02, which belonged to ST354.

Of the 42 *K. pneumoniae sensu stricto*, 22 distinct STs were identified, with 11 of them identified as CG14 and CG15 (ST14 and ST15, respectively). Only one was non-typeable (*K. pneumoniae* 0287-F3). Moreover, the majority of those belonging to CG14 and CG15 were the subpopulations of clonal food isolates (> 99.9% ANI). Eight isolates belonged to ST15, and three belonged to ST14, which differ only in their *infB* gene, but are however considered to be distinct CGs (Diancourt *et al.*, 2005; Rodrigues *et al.*, 2023). Both STs have been associated with extensive antibiotic resistance as prolific carbapenemase and ESBL producers, but little is known about their global expansion (Lam *et al.*, 2019; Turton *et al.*, 2019; Rodrigues *et al.*, 2023).

As expected, some of the *K. pneumoniae* food isolates that formed subpopulations of more than 99.9% ANI were identified as having the same ST and KL types as one another: of the eight ST15 *K. pneumoniae* genomes, five were capsular type KL24, and three were non-typeable. Additionally, two of the three *K. pneumoniae* ST14 genomes were capsular type KL16.

72

Figure 3.3. Core genome phylogeny of 73 *Klebsiella* genomes with summary MLST, virulence factor, AMR, defence system, and prophage data. Maximum-likelihood phylogeny of core gene alignment with ultrafast bootstrapping performed with IQ-TREE. Bootstrap support values at ≥ 70. Tree is midpoint rooted. Scale bar represents number of nucleotide substitutions per site. Tree rendering and heatmap were generated using iTOL. A total of 52 genomes isolated from different sources (coloured range) were included, with 21 reference strains (light grey), with presence (closed circles) or absence (open circles) from the physical collection. Isolates re-assigned to *Klebsiella* from *Raoultella* are indicated (blue box). Genomes were characterised by MLST, capsular type, and O-antigen loci. All genomes contained genes for acrAB efflux pump (dark orange). Gene predictions for oqxAB efflux pump are given as complete (yellow) or incomplete (light orange). Gene clusters for core virulence factor predictions are given as present (dark yellow-green) or incomplete (yellow-green). Accessory virulence factor predictions are given as present (dark pink), incomplete (pink), or absent (light pink). Gene clusters for flagella predictions are given as present (dark grey) or absent (grey). Heatmap (green) shows the predicted number of acquired AMR determinants according to antibiotic class and the number of plasmid variants. fim, type 1 fimbriae locus; mrk, type 3 fimbriae locus; ent, enterobactin locus; iuc, aerobactin locus; iro, salmochelin locus; ybt, yersiniabactin locus; clb, colibactin locus; rmpA/A2, regulators of mucoid phenotype genes; Chr, chromosomal.

Rodrigues *et al.* (2023) found that this capsular type constituted 14% of *K. pneumoniae* within CG14, while KL24 is prevalent in both clades. In addition to ST258, ST14, and ST15, ST101 and ST147 are also globally disseminated problem clones, often exhibiting XDR and PDR, neither of which were found among the *Klebsiella* spp. analysed (Turton *et al.*, 2019; Shankar *et al.*, 2021; Rodrigues *et al.*, 2023).

Four isolates of *Klebsiella* were used for phage hunting (**see Chapter 4**). *K. grimontii* KoN (and by extension its colony variant KoA) belonged to the ST215 and capsular type KL74. *K. grimontii* P038I ST and KL were both non-typeable. Whilst no ST or capsular type could be identified for *K. michiganensis* M7 21 2 #35 WT (hereafter *K. michiganensis* M7 WT) with high confidence, KL47 capsular type was predicted with low confidence, which could suggest the capsular loci spanned multiple contigs in the draft genome (Lam *et al.*, 2022). *K. pneumoniae* M26 18 2 #21 KpnN (and by extension its colony variant KpnA and parent strain M26 18 1) belonged to ST432 (CG1777), but capsule was non-typeable.

Although Kaptive/Kleborate are optimised for *K. pneumoniae* complex species, they identified six distinct STs and six capsular types among the *K. oxytoca* species complex (**Figure 3.3**, **Appendix 10.2**). The capsular type was non-typeable in 18 isolates, seven of which were in the *K. oxytoca* species complex, nine were in the *K. pneumoniae* species complex, and one was *K. ornithinolytica* 0220.AF01. Research has shown that specific capsular types are associated with hypervirulence, particularly in *K. pneumoniae sensu stricto* (Maatallah *et al.*, 2014; Choby *et al.*, 2020). Whilst some of the capsular types identified in the 62 *Klebsiella* genomes analysed here have been reported clinically, only one was associated with hypervirulence (Fung *et al.*, 2000; Corelli *et al.*, 2018; Ssekatawa

et al., 2021), capsular type KL54 identified in *K. variicola* DSM15968. Among *Klebsiella pneumoniae sensu stricto*, KL10, KL17, KL21, KL52, and KL62 were found in two isolates each (**Figure 3.3**, **Appendix 10.2**). KL16 (*n* = 4) and KL24 (*n* = 7) were the most prevalent capsular types in this strain collection, as described above.

The capsule of *Klebsiella* spp. is the best described virulence factor and is responsible for the mucoid phenotype that typifies the genus (Lawlor *et al.*, 2005; Pan *et al.*, 2015). The capsular polysaccharides can be used in typing, of which there are > 75 known types, with types K1 and K2 most commonly associated with clinical *K. pneumoniae* infections (Broberg *et al.*, 2014). Whole-genome analyses have superseded serotyping in identifying capsular types however, with the structural composition of novel capsule types being relatively unknown (Patro and Rathinavelan, 2019). The capsular polysaccharides that make up the capsule can be important primary phage receptors, but often undergo rapid genetic rearrangement to produce a distinct capsular type to evade immune recognition in the host, and/or phage predation (Westra *et al.*, 2015; Rendueles *et al.*, 2018; Holt *et al.*, 2020; Mangalea and Duerkop, 2020). Research has shown that capsular loss and/or serotype switching is common among *Klebsiella* spp., which can result in concomitant phage resistance (Haudiquet *et al.*, 2021; Haudiquet *et al.*, 2024). These data thus illustrate the importance of characterisation of clinical relevance of the bacterial hosts by determining their ST and their capsular types for rational design of phage preparations.

3.2.3 Identification of virulence determinants

The type strain *K. pneumoniae* DSM30104 was the only genome found to have the virulence factor *rmp2A* (**Figure 3.3**), which is an regulator of capsule production associated with increased capsule production and hypermucoidy in HvKP (Walker and Miller, 2020; Walker *et al.*, 2020). However, the presence of *rmpA* or *rmpA2* alone does not necessarily confer hypermucoidy. Indeed, Walker *et al.* (2020) found that increased capsule production and hypermucoidy may be separate traits dictated by the presence of other *rmp* genes such as *rmpD*.

An investigation of the O-antigen loci of the genomes in the collection showed a variety of types (**Figure 3.3**, **Appendix 10.2**). O1 variants (O1v1 or O1v2) were associated with 20 genomes and O2 variants (O2v1 or O2v2) were associated with ten genomes. Three O3/O3a and O5 variants were mainly associated with *K. quasipneumoniae*, *K. variicola* and *K. aerogenes* genomes. Other variants included OL101, OL102, and OL104, which were all associated with *K. pneumoniae* genomes. The most widely distributed O-antigen type among *K. pneumoniae* are O1 and O2 (Lam *et al.*, 2022), and is corroborated by the data presented here.

The O-antigen is a component of LPS found in the outer membrane of Gram-negative bacteria. O-antigen is a virulence factor that facilitates colonisation and inhibits

complement-mediated lysis during infection (Bengoechea and Sa Pessoa, 2019; Holt *et al.*, 2020), but O-antigens may also be important as phage receptors. LPS components are almost always identified as the secondary receptor in Gram-negative-targeting podoviruses (Molineux, 2006; Maffei *et al.*, 2021), with CPS components possibly primary receptors (Haudiquet *et al.*, 2024). Indeed, Tomás *et al.* (1987) identified three *Klebsiella* phages that may target the O-antigen component of the LPS as the primary phage receptor. Moreover, the O-antigen has been identified as the major receptor among the family *Autographiviridae* (Molineux, 2006; Maffei *et al.*, 2021).

Other virulence factors associated with *Klebsiella* infection include those associated with adherence, nutrient uptake/metabolism, and motility (**Figure 3.3**, **Appendix 10.2**). Many of these virulence factors are organised as gene clusters (loci) composed of multiple genes. The *fim* locus was found in all genomes but was incomplete in three foodborne *K. pneumoniae* and one *K. ornithinolytica*. The *mrk* locus was found complete in all but five genomes, where four were missing just one gene. *K. aerogenes* DSM30053 however, was missing most genes in the *mrk* locus.

Klebsiella spp. express two phenotypically-distinct types of fimbriae: type 1 fimbriae are encoded by the *fim* locus, whereas type 3 fimbriae are encoded by the *mrk* locus (Stahlhut *et al.*, 2012). Both are implicated in mucosal colonisation, biofilm formation, adhesion to medical devices, and immune activation, but are preferentially expressed depending on the niche environment (Stahlhut *et al.*, 2012). It is unclear whether the absence of specific genes in each locus would significantly affect the expression of either type of fimbriae or whether this might affect fitness and/or phage infection. Moreover, completeness of the genome assemblies still needs to be undertaken to determine whether missing genes are the result of a draft or incomplete genome.

The *ent* locus, which encodes the siderophore enterobactin, was found to be complete in all genomes, except *K. aerogenes* DSM30053, which was missing the *entD* gene (**Figure 3.3**, **Appendix 10.2**), but this strain did have all the genes associated with the salmochelin (*iro*) locus. As expected, presence of the yersiniabactin locus was variable: *ybt* was predicted in 19 of the 62 genomes, with 12 complete and seven incomplete (**Figure 3.3**, **Appendix 10.2**). No isolate other than DSM30053 had a complete salmochelin locus. In the majority of the other isolates, the salmochelin locus was missing *iroBCD*, but contained *iroE* and *iroN*. An exception was *K. ornithinolytica* 0220.ML01, which was missing only one gene (*iroC*) in the locus.

Classical strains of *K. pneumoniae* express the siderophore enterobactin, which is chromosomally-encoded, and occasionally yersiniabactin, which are encoded by the *ent* and *ybt* loci, respectively (Lawlor *et al.*, 2007; Walker and Miller, 2020). Additional siderophores aerobactin and salmochelin are usually associated with HvKP, and are encoded by *iuc* and *iro* loci, respectively (Walker and Miller, 2020). Whether the siderophore

loci require the presence of all genes for functional protein production is unclear however, thus, based on the genomes alone no conclusive inferences on the hypervirulence can be made.

Although *Klebsiella* spp. are not prolific toxin secretors like many other members of the *Enterobacteriaceae*, they can produce enterotoxins including tilimycin, tilivalline, and colibactin (Putze *et al.*, 2009; Hering *et al.*, 2019; Alexander *et al.*, 2020; Kienesberger *et al.*, 2022; Pöltl *et al.*, 2023). Tilivalline is a metabolic derivative of tilimycin, where tilimycin is encoded by ten genes (*npsA*, *npsB*, *thdA*, *aroX*, *dhbX*, *icmX*, *adsX*, *hmoX*, *mfsX*, and *uvrX*) within the *til* locus (Kienesberger *et al.*, 2022; Pöltl *et al.*, 2023). However, the locus is not included within the VFDB and presence of tilimycin among isolates within the collection was unclear. Only one strain, *K. pneumoniae* DSM30104, encoded some colibactin (*clb*) genes, with 13 out of 18 genes identified (**Figure 3.3**), suggesting the locus may be non-functional. The *clb* locus is usually associated with the *pks* island, which is a cluster of 18 *clb* genes that produce an array of toxigenic metabolites that are thought to facilitate inhibiting growth of other bacteria (Chagneau *et al.*, 2022). It is unclear whether the incomplete *clb* locus in *K. pneumoniae* DSM30104 was associated with the *pks* island, however. The incompleteness of the locus could again be due to the draft genome, and whether there is a functional *clb* locus is yet to be fully elucidated.

Genes necessary for allantoin metabolism were found complete in six genomes and incomplete in one genome (**Figure 3.3**, **Appendix 10.2**). Five of the genomes with complete allantoin locus were in the *K. oxytoca* species complex, with only one *K. pneumoniae*. Allantoin is a metabolic by-product of nucleic acid synthesis that can be utilised as an alternative nitrogen source, contributing to increased virulence in *Klebsiella* spp. (Martin and Bachman, 2018; Choby *et al.*, 2020). Indeed, allantoin metabolism is often associated with HvKP (Martin and Bachman, 2018).

K. aerogenes was re-assigned from the genus *Enterobacter* to the genus *Klebsiella*, and is the only motile member, producing peritrichous flagella as is typical for *Enterobacter* spp. Thus, it is unsurprising that flagella-associated proteins were found in this genome (**Figure 3.3**). Interestingly, a flagellar operon that appeared to be conserved in enteric bacteria such as *Yersinia* spp. and *E. coli* (Liu and Ochman, 2007) was identified in *K. quasipneumoniae* P057K W (**Figure 3.3**). One study found evidence for polar flagella in a clinical strain of *K. pneumoniae* (Carabarin-Lima *et al.*, 2016). Moreover, other studies have found that when grown under specific conditions, motility can be induced in other non-motile enteric bacteria (Carabarin-Lima *et al.*, 2016).

The type VI secretion system (T6SS) was found in all *Klebsiella* isolates (data not shown). Bacterial secretion systems are fundamental to bacterial survival, with the T6SS being the main secretion system in *Klebsiella* spp. that facilitates their ability to occupy a particular niche. *Klebsiella* species are distinct from most other Enterobacterales in that they do not

encode a type III secretion system, and the type IV secretion system is highly variable among species, usually being associated with large conjugative plasmids (Zheng *et al.*, 2020; Low *et al.*, 2022).

3.2.4 Identification of AMR determinants

Most *Klebsiella* spp. harbour chromosomally-encoded β-lactamase genes (*bla*), and many strains can have additional variants encoded extra-chromosomally, usually on plasmids (summarised in **Table 1.2**). As expected, almost all of the genomes were found to have a chromosomally-encoded β-lactamase gene (**Figure 3.3**, **Appendix 10.2**). One *K. pneumoniae* genome was found to have a truncated $bla_{SHV-120}$ (**Appendix 10.2**), whereas no β-lactamase was found in another *K. pneumoniae* isolate or *K. aerogenes* (**Figure 3.3**).

bla_{OXY} variants encode OXY enzymes (penicillinases) and appear to have evolved at a similar rate to MLST housekeeping genes within the *K. oxytoca* species complex (Fevre *et al.*, 2005). All bla_{OXY} variants within the *K. oxytoca* species complex did indeed correspond to their phylogroup (**Table 1.1**, **Figure 3.3**). Specifically, *K. michiganensis* belongs to phylogroup 1 and harboured bla_{OXY-1}; *K. huaxiensis* belongs to phylogroup 3 and harboured bla_{OXY-3}; *K. pasteurii* belongs to phylogroup 4 and harboured bla_{OXY-4}; and *K. grimontii* belongs to phylogroup 6 and harboured bla_{OXY-6}. Additionally, *K. michiganensis* 0220.ML04 was predicted to encode bla_{OXA-10} and bla_{GES-5}. The bla_{OXA-10} gene encodes OXA-10 (class D β-lactamase) which is not considered a carbapenemase but does have ESBL activity (Queenan and Bush, 2007; Maurya *et al.*, 2017). bla_{GES-5} encodes GES-5, a (class A) carbapenemase (Queenan and Bush, 2007; Tooke *et al.*, 2019; Sawa *et al.*, 2020). Importantly, GES enzymes are inhibited by the β-lactamase inhibitor clavulanic acid, whereas OXA enzymes can show variable inhibition depending on the variant (**Table 1.2**) (Queenan and Bush, 2007; Tooke *et al.*, 2019; Sawa *et al.*, 2020).

Chromosomally-encoded bla_{SHV-1} and bla_{SHV-11} variants were predicted in most *K. pneumoniae* genomes, whose encoded enzymes SHV-1 and SHV-11 are not considered true ESBLs as they only hydrolyse early cephalosporins (**Figure 3.3**) (Hæggman *et al.*, 2004; Tooke *et al.*, 2019). Additional variants of bla_{SHV} were found among *K. pneumoniae* genomes, including bla_{SHV-27}, bla_{SHV-28}, bla_{SHV-33}, bla_{SHV-36}, bla_{SHV-61}, $bla_{SHV-119}$, and $bla_{SHV-189}$ (**Figure 3.3**), whose enzymes are associated with extended-spectrum activity (Tooke *et al.*, 2019; Sawa *et al.*, 2020). The bla_{OKP-A2} gene was found in *K. quasipneumoniae* P057K W and its encoded enzyme is also not considered an ESBL, having only variable activity against early cephalosporins (Hæggman *et al.*, 2004; Fevre *et al.*, 2005). Four different bla_{LEN} variants were identified among the four *K. variicola* genomes. Similar to the bla_{OXY} variants in the *K. oxytoca* species complex, bla_{SHV}, bla_{OKP}, and bla_{LEN} variants appear to be associated with specific species, with bla_{OKP} and bla_{LEN} able to reliably distinguish *K. quasipneumoniae* and *K. variicola* from other *K. pneumoniae* species complex species (Rocha *et al.*, 2022).

The bla_{PLA1a} gene was predicted in the two *K. ornithinolytica* genomes (**Figure 3.3**) and appears to be a β-lactamase specific to these species, as well as *K. planticola* (Walckenaer *et al.*, 2004).

The fosfomycin resistance gene, *fosA* was predicted in 60 of the 62 isolates (**Figure 3.3**, **Appendix 10.2**), and is usually chromosomally-encoded in *Klebsiella* spp. (Thompson *et al.*, 2015; Byarugaba *et al.*, 2023). Additionally, *oqxAB*, which encodes the OqxAB efflux pump was predicted in all genomes except *K. pasteurii*. The *oqxAB* genes can be chromosomally-encoded or plasmid-mediated (Li *et al.*, 2019), whereas *acrAB*, which encodes the AcrAB efflux pump is almost exclusively chromosomally-encoded (Lv *et al.*, 2020) and was found in all genomes.

A macrolide resistance gene (*mphA*) was only present in *K. pneumoniae* NCTC13438 (ST258) and a quinolone resistance gene (*qnrS1*) was found in only two foodborne *K. pneumoniae* genomes (**Figure 3.3**, **Appendix 10.2**). Genes encoding resistance to phenicols (encoded by *cmlA1*, *catA1*, *florR*), sulfonamides (encoded by *sul1*, *sul2*, *sul3*) and trimethroprim (encoded by *dfrA*) was only predicted in a subset of genomes however, where all were classified as MDR isolates (**Figure 3.3**, **Appendix 10.2**).

Three *K. michiganensis* genomes were found to have one or both *aph(3')-Ia* and *ant(2")-Ia* genes predicted, which encode two aminoglycoside modifying enzymes (AMEs). Aminoglycoside resistance was more commonly associated with MDR genomes, except for *K. michiganensis* 0220.ML04 (**Figure 3.3**). This isolate harboured two aminoglycoside resistance genes, and three β-lactamase genes, but was not predicted MDR. AMEs include acetyltransferases encoded by the *aac* gene; nucleotidyltransferases encoded by *ant* genes; or phosphotransferases encoded by *aph* genes (Ramirez and Tolmasky, 2010). The number of genes and gene variants encoding AMEs is extensive and largely attributed to rapid genetic alterations as well as ready mobilisation and transfer between bacteria (Ramirez and Tolmasky, 2010). The accumulation of multiple genes and variants encoding enzymes with different spectrums of activity within a single bacterial isolate is therefore not uncommon (Ramirez and Tolmasky, 2010). A comprehensive review of AMEs, their nomenclature, mechanism of action, and spectrum of activity is available from Ramirez and Tolmasky (2010).

Interestingly, the two *K. pneumoniae* ST38 strains each had three extra-chromosomal predicted *bla* genes; bla_{SHV}, bla_{OXA}, and bla_{TEM}, where the latter encodes a TEM-1 penicillinase, but the other genes would typically encode ESBLs (**Table 1.2**). These two strains also harboured seven AME genes (including one or more *aac(6')*, *ant(2")*, *ant(3")*, *aph(3')*, *aph(3")*, *aph(6)*, and variants) (**Figure 3.3**). Moreover, these two isolates both had predicted genes whose enzymes can confer resistance to chloramphenicol (*cmlA1* and *catA1*), fosfomycin (*fosA*), sulfamethoxazole (*sul1* and *sul2*), and tetracycline (*tetD*), along with genes for AcrAB and OqxAB MDR efflux pumps. These data suggest that the two *K.*

pneumoniae ST38 strains demonstrate potential resistance to at least seven of the nine antibiotic classes and are therefore XDR. A further 14 *K. pneumoniae* isolates were identified as putative MDR, with resistance to at least three different classes of antibiotics (**Figure 3.3**, **Appendix 10.2**). Moreover, a further eight *K. pneumoniae* isolates were identified as putative XDR, with resistance to all except two classes of antibiotics (inclusive of those conferred by efflux pumps) (**Figure 3.3**, **Appendix 10.2**).

The number of predicted plasmids per genome was highly variable, varying between zero and six (mean = 2.5, median = 2, ± SD = 1.72) (**Figure 3.3**, **Appendix 10.2**). Whilst no correlation was found between plasmid number and isolates that were both MDR (*n* = 17) and XDR (*n* = 8), there was a slight difference between those that were XDR only (mean = 3.6, ± SD = 1.22) compared to all isolates (*n* = 62).

Several clinically relevant plasmids were predicted in multiple isolates, including IncF, IncN, and IncR types (**Figure 3.3**, **Appendix 10.2**), which have been implicated in conferring MDR in *Enterobacteriaceae* (Rozwandowicz *et al.*, 2018; Zaman *et al.*, 2018; Stoesser *et al.*, 2020). Additionally, several smaller plasmids were also predicted in multiple isolates, such as Col440. The two *K. pneumoniae* ST38 and *K. pneumoniae* P008E 1 were predicted to harbour IncFIB(pQIL), which has been associated with carbapenem-resistant *K. pneumoniae* infections (Stoesser *et al.*, 2020).

Interestingly, many of the isolates that were clonal (> 99.9% ANI) exhibited differences in their composition of plasmid and AMR determinants, defence systems, and prophages (**Figure 3.3**, **Appendix 10.2**). However, most are draft genomes, and *in silico* analyses may show differences in their predictions if certain genes span more than one contig or are missing from the assembly.

The genome analyses of the *Klebsiella* isolates described show a large amount of data relating to their sequence type, capsular and O-antigen type, and their virulence and AMR profiles, having revealed considerable diversity. These data will be useful to understanding phage predation, particularly with antibiotic susceptibility testing and kinetics assays.

3.2.5 Prophage composition and anti-phage defence mechanisms

Defence systems were widely distributed among all the genomes analysed, ranging from between 3 and 16 (mean = 8.94, median = 9, ± SD = 3.38) (**Figure 3.3**, **Appendix 10.2**). There was no clear correlation between defence system and the number of prophages predicted. There was also no clear correlation between the number of prophages and the number of AMR/plasmids predicted for all isolates. However, the mean number of prophages predicted among all isolates (mean = 3.88, ± SD = 1.92) was slightly higher than the mean prophages for those considered XDR only (mean = 3.13, ± SD = 1.83).

3.3 Summary of the diversity of *Klebsiella* isolates in the collection

A total of 52 *Klebsiella* spp. were isolated from a range of different sources or derived from previous studies (**Table 3.1**, **Figure 3.4A-D**). Of the 52 isolates, 43 genomes were analysed, leaving nine in the collection that have not been sequenced or analysed, but were included in some of the phage phenotyping (all clinical strains with the code "CL" (under a Materials Transfer Agreement preventing sequencing), and *K. michiganensis* M7 21 2 #35 Oda^R (hereafter *K. michiganensis* M7 Oda^R). A further 19 *Klebsiella* genomes isolated from food samples were analysed, but were not added to the physical isolate collection as they are available in other collections within QIB, of which the majority were *K. pneumoniae* (**Figure 3.4B**). This gave a total of 71 isolates, of which two were from an unknown source and excluded from **Figure 3.4**. Additionally, another 21 reference genomes were downloaded from the NCBI database and were included. One genome, a presumptive *Klebsiella* isolate could not be assembled due to contamination and was excluded.
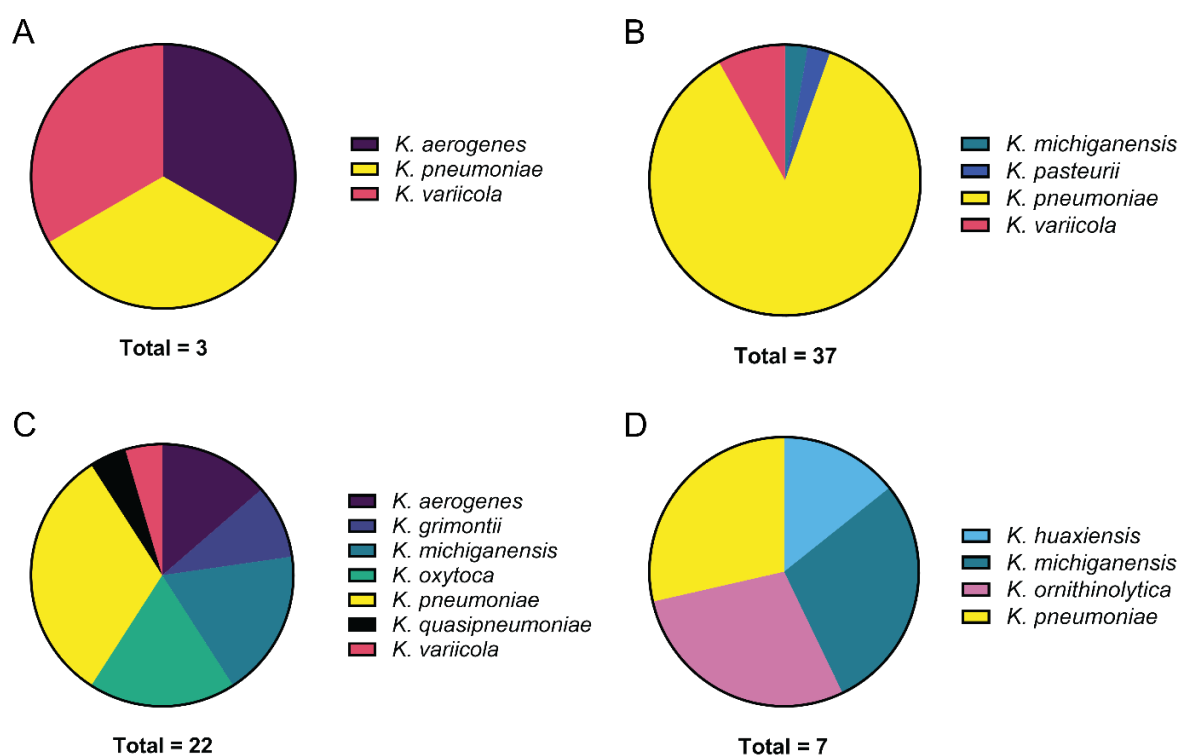
Figure 3.4. Diversity of 69 *Klebsiella* spp. isolated from different sources. *Klebsiella* isolates assorted by species among **(A)** type strains; **(B)** food strains; **(C)** clinical strains; and **(D)** wastewater strains. Pie size corresponds to the number of isolates with the total number of isolates from the source given below. A total of 69 isolates were included, with two isolates from an unknown source (*K. pneumoniae* ST38 variants, data not shown). Two clinical *K. oxytoca* isolates were reclassified as *K. grimontii*.

Among all the isolates, the majority of food (*n* = 32) and clinical (*n* = 7) isolates were *K. pneumoniae*. Most diversity was found within the clinical samples, with equal number of *K. michiganensis* (*n* = 4) and *K. oxytoca* (*n* = 4). However, one of the *K. oxytoca* was a single colony variant of the other (*K. oxytoca* M59 22 8 KoA and *K. oxytoca* M59 22 8 KoN, respectively). Moreover, it was later discovered through ANI analyses that two of the *K. oxytoca* strains were in fact *K. grimontii* and were renamed *K. grimontii* M59 22 9 KoN and *K. grimontii* M59 22 9 KoN (see ANI analyses in **section 3.2.1**). Wastewater contained isolates that are less commonly associated with human infection, including *K. huaxiensis* (*n* = 1) and *K. ornithinolytica* (*n* = 2). *K. michiganensis* was the second most commonly isolated species after *K. pneumoniae* in three of the four sources. Several species were missing from the collection, including some former *Raoultella* spp., *K. spallanzanii*, and *K. oxytoca*. Whilst the clinical relevance of some species remains unclear, *K. oxytoca* is the second most commonly-isolated species after *K. pneumoniae* within hospital settings (Broberg *et al.*, 2014) and would be useful to have as part of the collection.

3.4 Conclusions

The diversity of *Klebsiella* isolates was explored to investigate the biocontrol potential of the phages isolated during this project. Together, a total of 52 *Klebsiella* isolates spanning almost all species were analysed, with an additional 19 foodborne isolates whose genomes have been analysed to determine which to add to the isolate collection, based on the species clonality, and the diversity of their ST, capsular and O-antigen loci.

ANI, pangenome, and phylogeny data supported correct taxonomic assignment for all species, except for two *K. oxytoca* isolates, which were re-assigned to *K. grimontii*. The pangenome describes the composition of core and accessory genes among a group of organisms (Vernikos *et al.,* 2015; Sitto and Battistuzzi, 2019). It can be used in speciation of closely-related organisms, to understand the promiscuity, and/or track the evolutionary history of particular traits (Sitto and Battistuzzi, 2019).

Many of the *in silico* analyses found factors associated with virulence, and in one case hypervirulence, and AMR. These data provide an insight on clinical relevance of the *Klebsiella* hosts, which will inform phage host range assays, kinetics, and future phage preparation compositions. *K. pneumoniae* and *K. oxytoca* are clinically overrepresented, likely due to inadequate speciation due to limited resources within hospital settings (Fontana *et al.*, 2019). For example, infections caused by members of the *K. pneumoniae* species complex cannot be distinguished from one another (Fontana *et al.*, 2019). This means epidemiology of certain species such as *K. quasipneumoniae* is not well understood, and likely underestimated. Moreover, *K. electrica*, *K. ornithinolytica*, *K. planticola*, and *K. terrigena* are all relatively understudied (Appel *et al.*, 2021).

The draft status of some of the genomes analysed here could have led to inaccurate ST, capsular type, virulence factor, and AMR gene predictions. However, improved accuracy could be achieved through long-read-only assembly of bacterial genomes, together with long-read and short-read polishing steps (Wick *et al.*, 2021; Wick *et al.*, 2023), which was beyond the scope of this project.

Together, these data will prove useful in understanding the evolution and adaptation of the hosts in response to phage infection. Particularly important was understanding the genomic background of the parent isolate of the TraDIS mutant library, *K. michiganensis* M7 WT.

CHAPTER 4: PHAGE DISCOVERY AND PHENOTYPIC CHARACTERISATION

4.1 Introduction

Phages are the most abundant and diverse biological entities on Earth (Hendrix, 2002; Mushegian, 2020). They occupy almost all microbial niches, where they play an essential role in the ecology and evolution of their bacterial hosts (Mavrich and Hatfull, 2017; Ofir and Sorek, 2018; Chiang *et al.*, 2019; Al-Shayeb *et al.*, 2020; Dion *et al.*, 2020; Mushegian, 2020).

Historically, phage classification has been based on nucleic acid composition and morphology (Ackermann, 2005). Whilst the nucleic acid composition of phages can be highly diverse, tailed (dsDNA) phages are overrepresented in phage isolation (Callanan *et al.*, 2018). In addition to their genomic and morphological diversity, phages can also adopt a diverse array of life cycles, of which, the lytic and lysogenic life cycles are the most well-characterised (Mäntynen *et al.*, 2021).

In vitro characterisation of phages is essential to understanding infection biology and to inform *in vivo* studies and determine clinical value for use as therapeutics (Abedon, 2017). For example, the ability of temperate phages to transfer genetic material between hosts and their integration into the host genome without immediate lysis makes them unsuitable in phage therapy settings (Hay and Lithgow, 2019; Romero-Calle *et al.*, 2019). Moreover, characterisation of phage infection dynamics in a particular host or a panel of hosts is integral to maximising efficacy and mitigating any potential antagonism when phages are combined with either antibiotics and/or other phages (Romero-Calle *et al.*, 2019). Indeed, previous clinical applications have failed due to reduced stability and low titre of phage cocktails (Jault *et al.*, 2019).

Phage resistance is a phenomenon driven by selective pressure that phages exert on their bacterial hosts. It can be argued that phage resistance is a limitation of phage therapy, but can also be advantageous in that it may result in a concomitant reduction in bacterial virulence and/or antibiotic resistance (Hesse *et al.*, 2020). Predicting the type of resistance emergence and subsequent impact on bacterial fitness is difficult, however.

The co-evolution of bacteria in response to phage infection is a major regulatory hurdle. Improved characterisation of phages will inform the type of bacterial receptors used during infection, the nature of changes in bacterial virulence, and the nature of resistance emergence when using phages as single therapeutics, in combination with antibiotics, and/or other phages. Together, these data will build on existing knowledge of phage infection dynamics so therapeutics can be rationally designed and personalised to the individual and/or bacterial host(s).

This chapter will examine phage discovery and characterisation with results and discussion combined for each section.

4.1.1 Aims

1. Isolate and build a collection of phages against a panel of diverse *Klebsiella* spp.
2. Characterise phage host range and infection dynamics
3. Determine phage-antibiotic synergy with a range of clinically relevant antibiotics

4.2 Phage isolation

Phage isolation involved the collection of samples, enrichment, single-plaque purification, amplification, as well as DNA extraction and sequencing to distinguish the phage isolates from each other (**Figure 4.1**). Enrichment entailed incubating potential phages in a collected sample with the target bacterial host. Single-plaque purification of phages was performed by spot testing enrichments after which positive samples were titred and individual plaques were propagated by plaque assay to distinguish them from clearance without individual plaques, a phenomenon known as lysis-from-without (Abedon, 2011). Productive infection was defined as phages that could produce individual plaques. Clearance zones on the agar without plaques at lower dilutions were not counted as productive infection and were considered lysis-from-without (Abedon, 2011). Where multiple plaque morphologies occurred in the same sample, each distinct morphology was propagated from a single plaque and amplified. Once phages were amplified, they were centrifuged and filtered to produce a crude single-phage lysate. From the lysate, further downstream analyses could be performed, and phage DNA was extracted and sequenced for preliminary identification.
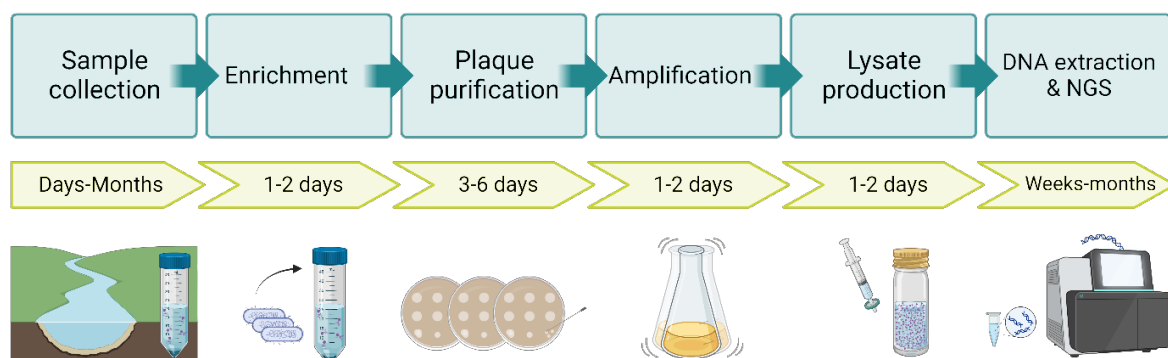
Figure 4.1. Schematic representation of phage isolation timeline. Phage isolation starts with sample collection, which can take days to months. The sample is then processed to enrich phages specific for the bacteria of interest and spot tested by double agar overlay. Positive spots are titred and single-plaque morphologies purified. Once purified, a single plaque is amplified by inoculating with the bacteria of interest, usually the same one against which it was enriched. Following small-scale amplification, the lysate is centrifuged and filtered to separate it from bacterial cells and other large debris into a crude lysate. The process of enrichment to lysate production can take one to two weeks. From the lysate, further assays can be conducted, such as phage infection assays and phages can be reamplified from the lysate. Following initial lysate production, the phage DNA is extracted and sequenced.

Thirty-five virulent phage isolates were obtained from wastewater and river water samples using six different *Klebsiella* isolation hosts spanning five different species (**Appendix 10.3**). Of these, phages isolated from the same sample that exhibited different plaque morphologies were purified out and sequenced. After sequencing 26 genetically distinct phages were observed, that is, phages with < 100% sequence similarity across the genome length (**Table 4.1**). A more in-depth description and discussion of the phage genomics is presented in **Chapter 6**. Following assembly, phages were preliminarily assigned a genus according to their closest database relative, according to the currently established ICTV genus demarcation criterion of 70% nucleotide sequence similarity over the genome length (see **Chapter 6**) (Turner *et al.*, 2021). The phages within the newly established phage collection covered five known genera and two new genera within four families: *Slopekvirus* and *Jiaodavirus* (both in the family *Straboviridae*); *Przondovirus* (in the family *Autographiviridae*); *Webervirus* (in the family *Drexlerviridae*); two new genera of unclassified siphoviruses and a new species of unclassified myovirus in the class *Caudoviricetes* (**Table 4.1**).

Initially, two clinical *Klebsiella* isolates from two different species were used as hosts for the enrichment process, *K. michiganensis* M7 WT and *K. pneumoniae* KpnN (**Table 4.1**, **Appendix 10.3**). Following isolation of 12 phages against these two clinical isolates, a further four *Klebsiella* isolates from three different species were used as hosts to isolate the other 13 phages, including two different clinical *K. grimontii* strains, a clinical *K. quasipneumoniae* strain, and a type strain of *K. variicola* (**Table 4.1**, **Appendix 10.3**).

Klebsiella phages Bolond and Fifoon were isolated from the same river water sample against *K. michiganensis* M7 WT and *K. pneumoniae* KpnN, respectively (**Appendix 10.3**). *Klebsiella* phage Oda was isolated from a different river water sample against *K. michiganensis* M7 WT (**Appendix 10.3**). *Klebsiella* phages Assynt, Torridon, and Suilven were isolated from different processes of the same wastewater plant, whereas Blaven was isolated from a different river water sample, but genetically similar to the former three phages (see **Chapter 6** and **Appendix 10.3**). All were isolated using *K. pneumoniae* KpnN as the host strain. Initially, only a single-plaque morphology was found in the enrichment of *Klebsiella* phage Oda. However, after using the enrichment again for training purposes, another *Klebsiella* phage, SteelHaze was isolated and purified based on distinct plaque morphology. All other phages were isolated from three different wastewater plants (**Appendix 10.3**).

Table 4.1. Names and lab codes of 26 distinct virulent phages isolated using six different strains of *Klebsiella* spp. and their preliminary taxonomic assignment.

Phage name	Lab code	Source	Isolation host	Preliminary taxonomic assignment*
Klebsiella phage Bolond	CE1	River Bure, Norwich	*K. michiganensis* M7 21 2 #35 WT	*Straboviridae, Slopekvirus*
Klebsiella phage Fifoon	CE2	River Bure, Norwich	*K. pneumoniae* M26 18 2 #21 KpnN	*Straboviridae, Slopekvirus*
Klebsiella phage Oda	CE3	River Yare, Norwich	*K. michiganensis* M7 21 2 #35 WT	*Autographiviridae, Studiervirinae, Przondovirus*
Klebsiella phage Assynt	CE4	Wastewater A1	*K. pneumoniae* M26 18 2 #21 KpnN	*Straboviridae, Tevenvirinae, Jiaodavirus*
Klebsiella phage Torridon	CE5	Wastewater A2	*K. pneumoniae* M26 18 2 #21 KpnN	*Straboviridae, Tevenvirinae, Jiaodavirus*
Klebsiella phage Suilven	CE6	Wastewater A3	*K. pneumoniae* M26 18 2 #21 KpnN	*Straboviridae, Tevenvirinae, Jiaodavirus*
Klebsiella phage Blaven	CE7	River Tud, Norwich	*K. pneumoniae* M26 18 2 #21 KpnN	*Straboviridae, Tevenvirinae, Jiaodavirus*
Klebsiella phage Toyotomi	CE8	Wastewater A1	*K. michiganensis* M7 21 2 #35 WT	*Autographiviridae, Studiervirinae, Przondovirus*
Klebsiella phage Mera	CE9	Wastewater A2	*K. michiganensis* M7 21 2 #35 WT	*Autographiviridae, Studiervirinae, Przondovirus*
Klebsiella phage Speegle	CE10	Wastewater A3	*K. michiganensis* M7 21 2 #35 WT	*Autographiviridae, Studiervirinae, Przondovirus*
Klebsiella phage Cornelius	CE11	Wastewater A1	*K. michiganensis* M7 21 2 #35 WT	*Autographiviridae, Studiervirinae, Przondovirus*
Klebsiella phage Tokugawa	CE12	Wastewater A2	*K. michiganensis* M7 21 2 #35 WT	*Autographiviridae, Studiervirinae, Przondovirus*
Klebsiella phage SteelHaze	CE14	River Yare, Norwich	*K. michiganensis* M7 21 2 #35 WT	Unclassified myovirus
Klebsiella phage Saitama	CE15	Wastewater B	*K. quasipneumoniae* P057K W	*Autographiviridae, Studiervirinae, Przondovirus*
Klebsiella phage Iolaire	CE18	Wastewater C	*K. quasipneumoniae* P057K W	Unclassified siphovirus
Klebsiella phage Fioreun	CE19	Wastewater C	*K. quasipneumoniae* P057K W	Unclassified siphovirus
Klebsiella phage Emom	CE20	Wastewater B	*K. grimontii* M59 22 8 KoN[a]	*Autographiviridae, Studiervirinae, Przondovirus*
Klebsiella phage Amrap	CE22	Wastewater C	*K. grimontii* M59 22 8 KoN[a]	*Autographiviridae, Studiervirinae, Przondovirus*
Klebsiella phage Kandur	CE24	Wastewater B	*K. grimontii* P038I	*Straboviridae, Tevenvirinae, Jiaodavirus*
Klebsiella phage NejNej	CE25	Wastewater B	*K. grimontii* P038I	*Straboviridae, Tevenvirinae, Jiaodavirus*
Klebsiella phage Pigeon	CE26	Wastewater C	*K. grimontii* P038I	*Straboviridae, Tevenvirinae, Jiaodavirus*
Klebsiella phage Sidders-16	CE27	Wastewater C	*K. grimontii* P038I	*Straboviridae, Tevenvirinae, Jiaodavirus*

Klebsiella phage Eggie	CE28	Wastewater D	*K. grimontii* P038I	*Straboviridae*, *Tevenvirinae*, *Jiaodavirus*
Klebsiella phage Whistle	CE30	Wastewater B	*K. variicola* DSM15968	*Autographiviridae*, *Studiervirinae*, *Przondovirus*
Klebsiella phage DeeWhy	CE32	Wastewater C	*K. variicola* DSM15968	*Drexlerviridae*, *Webervirus*
Klebsiella phage Dereham	CE34	Wastewater D	*K. variicola* DSM15968	Unclassified siphovirus

Wastewater A, wastewater treatment plant located in SE England; A1, raw sewage; A2, aeration feed; A3, mixed liquor; Wastewater B-D, three different wastewater treatment plants located in NW England; *, taxonomic assignment given as family, subfamily (if assigned), and genus. See **Chapter 6** for in-depth taxonomic analyses; [a], previously misassigned as *K. oxytoca* M59 22 8 KoN. See **Chapter 3** for analyses.

4.3 Phage morphology

One or two phages representative of each genus were imaged by transmission electron microscopy (TEM), performed by QIB Advanced Microscopy Facility and JIC Bioimaging team members. The virion sizes were measured on the raw images using Omero software. Measurements were taken for a minimum of eight different particles for capsids and tails for each phage in an image, and the means and standard deviation (± SD) calculated (**Table 4.2**). Phage DeeWhy (*Webervirus*) was representative of the family *Drexlerviridae*; phages Bolond (*Slopekvirus*) and Suilven (*Jiaodavirus*) were representative of the family *Straboviridae*; phages Oda and Whistle (*Przondovirus*) were representative of the family *Autographiviridae*; phages SteelHaze and Dereham were representative of unclassified myovirus and siphovirus, respectively. The other unclassified siphoviruses Iolaire and Fioreun, which were genomically distinct to Dereham, were not imaged. Their morphology was inferred from the genomic characterisation.

Table 4.2. Mean measurements of phage capsid and tail sheath in nm, with ± SD values. Measurements of empty capsids and contracted tails were excluded.

Phage (*n*)	Capsid measurements nm (± SD)		Tail measurements nm (± SD)	
	Length	**Width**	**Length**	**Width**
Drexlerviridae				
DeeWhy (*10*)	66.7 (*1.64*)	65.4 (*1.27*)	161.8 (*1.63*)	9.9 (*0.69*)
Straboviridae				
Bolond (*29*)	117.5 (*3.35*)	86.2 (*3.51*)	111.9 (*2.83*)	22.5 (*2.36*)
Suilven (*18*)	116.5 (*2.64*)	83.8 (*3.53*)	108.9 (*3.50*)	24.8 (*2.40*)
Autographiviridae				
Oda (*16*[a])	60.9 (*1.98*)	-	19.4 (*2.46*)	-
Whistle (*16*[b])	60.9 (*1.25*)	56.3 (*1.83*)	18.0 (*2.88*)	-
Unclassified myoviruses				
SteelHaze (*28*[c])	70.7 (*2.27*)	63.13 (*1.83*)	81.4 (*3.90*)	18.7 (*1.18*)
Unclassified siphoviruses				
Dereham (*12*)	63.5 (*3.01*)	60.4 (*2.08*)	106.9 (*3.51*)	9.5 (*0.73*)

[a], capsids *n* = 16, tails *n* = 5; [b], capsids *n* = 16, tails *n* = 8; [c], capsids *n* = 28, tails *n* = 17; -, not measured.

Myoviruses in the family *Straboviridae* included phages in the genera *Slopekvirus* (**Figure 4.2A**) and *Jiaodavirus* (**Figure 4.2B**), and had large, prolate icosahedral capsids and contractile tails making them ~227 nm in total length. The unclassified myovirus *Klebsiella* phage SteelHaze (**Figure 4.2C**) had a different morphology to the other myoviruses in that it had an isometric icosahedral capsid (~71 nm) and a contractile tail (~81 nm). Interestingly, in some images, SteelHaze appeared to have a different tail sheath and baseplate structure

to the phages in the family *Straboviridae* further confirming that they should not be assigned to the same family.
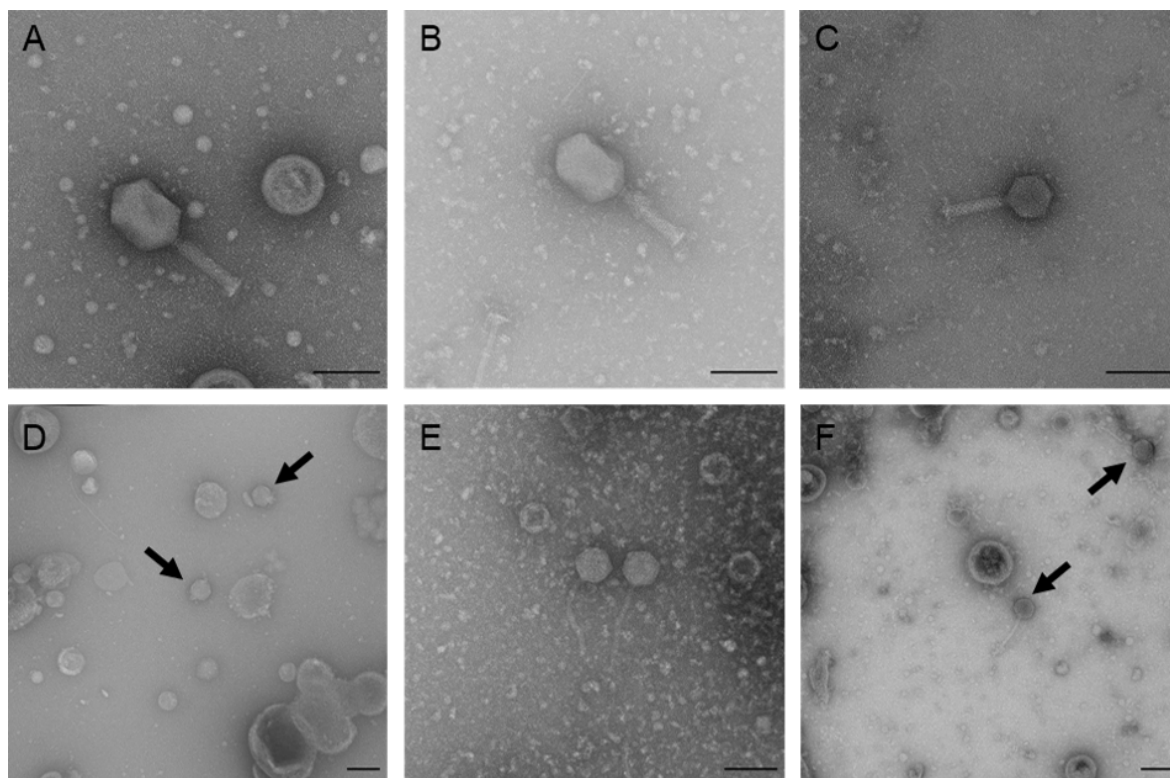
Figure 4.2. Transmission electron microscopy images of phages representative of each genus. (A) *Klebsiella* phage Bolond, a myovirus in the genus *Slopekvirus*. **(B)** *Klebsiella* phage Suilven, a myovirus in the genus *Jiaodavirus*. **(C)** *Klebsiella* phage SteelHaze, an unclassified myovirus in the class *Caudoviricetes*. **(D)** *Klebsiella* phage Oda (arrows), a podovirus in the genus *Przondovirus*. **(E)** *Klebsiella* phage DeeWhy, a siphovirus in the genus *Webervirus*. **(F)** *Klebsiella* phage Dereham (arrows), an unclassified siphovirus in the class *Caudoviricetes*. Scale bars represent 100 nm.

Myoviruses have complex tail structures, for example, phage T4 (representative of the family *Straboviridae*) tail sheath consists of multiple copies of gene product (gp)19, which appear like concentric rings that surround the inner tail tube (Leiman *et al.*, 2004). Tail sheaths of both Bolond (**Figure 4.2A**) and Suilven (**Figure 4.2B**) exhibited similar structures to T4. The structure of the tail sheath of the unclassified myovirus SteelHaze however, appeared to be different (**Figure 4.2C**), showing similarities to phages whose tail sheaths contain helical arrangements of tail sheath proteins (Yang *et al.*, 2023).

Phages in the family *Autographiviridae* were morphologically classified as podoviruses (**Figure 4.2D**), with small icosahedral capsids (~61 nm) and short, non-contractile tails (~19 nm), many of which were too small to distinguish from debris in the TEM sample. The phages in the family *Drexlerviridae* (**Figure 4.2E**) and unclassified siphovirus *Klebsiella* phage Dereham (**Figure 4.2F**) showed siphovirus morphology, with isometric icosahedral capsids (~65 nm and ~62 nm, respectively) and long, flexible, non-contractile tails. Their tail lengths differed by a considerable margin however (~160 nm and ~107 nm, respectively).

4.4 Phage infection dynamics

4.4.1 Phage host range determination

To examine the host range, the 26 genomically distinct phages were tested by plaque assay against a collection of *Klebsiella* from different species and sources, and represented a range of capsule and sequence types (**Figure 4.3**). Phages were tested against a minimum of 34 and a maximum of 38 *Klebsiella* isolates, with an average number of isolates able to be productively infected across all phages at 6.96 (± SD = 6.88) (**Appendix 10.4**).

Phages from specific genera exhibited differences in their host range, being categorised as having broad or narrow host range. Since no standardised definition exists, broad host range was arbitrarily defined as phages that could infect at least four different *Klebsiella* species (Ross *et al.*, 2016). Generally, phages with larger genomes (> 50 kbp) and myovirus morphotype exhibited broader host range than those with smaller genomes (**Figure 4.3**).

Phages preliminarily assigned to the genus *Przondovirus* exhibited very narrow host range, with eight capable of infecting only a single *Klebsiella* isolate (their isolation host) of the 38 isolates tested (2.6%), regardless of species, or capsular type (**Figure 4.3**, **Appendix 10.4**). None of the przondoviruses were capable of infecting the phage Oda resistant mutant, *K. michiganensis* M7 Oda^R, suggesting a mechanism of cross-resistance. Two przondoviruses, Emom and Amrap were capable of productively infecting two clinical strains (5.3%), their isolation host *K. grimontii* KoN and *K. oxytoca* CL4. Interestingly, both phages were not capable of infecting the SCV of their isolation host *K. grimontii* KoA. Nine of the ten przondoviruses did not exhibit any lysis without productive infection against any of the tested isolates. *Klebsiella* phage Whistle did exhibit lysis without productive infection when tested against three *K. pneumoniae* isolates, two of which had capsular types KL16 (*K. pneumoniae* H-27-MC778-782col38B) and KL30 (*K. pneumoniae* 0220.RS01) and one non-typable (*K. pneumoniae* KpnN). Indeed, Whistle was the only phage capable of causing lysis without productive infection against *K. pneumoniae* 0220.RS01. The przondoviruses have a relatively small genome of approximately 42 kbp (see **Chapter 6**), and this may limit their host expansion capabilities, including their ability to acquire multiple depolymerases.

The two slopekviruses, Bolond and Fifoon exhibited broad host range (**Figure 4.3**, **Appendix 10.4**). Bolond was capable of infecting fewer *Klebsiella* isolates than Fifoon, at 11 and 15, respectively of a total 36 tested (30.6% and 41.7% of isolates, respectively). Generally, both phages were unable to productively infect or lyse from without the same strains, except for two clinical *K. aerogenes* strains, against which Fifoon was able cause lysis without productive infection. There was no correlation between their ability or inability to infect across isolate sources, ST, or capsular type. Interestingly, both Bolond and Fifoon were unable to productively infect the WT *K. grimontii* KoN, but were able to infect the SCV *K. grimontii* KoA, which is in contrast to the przondoviruses Emom and Amrap. However,

only Fifoon was capable of productively infecting both the WT *K. pneumoniae* KpnN and the SCV *K. pneumoniae* KpnA. This could be because the WT *K. pneumoniae* KpnN was the isolation host for Fifoon, but not Bolond.
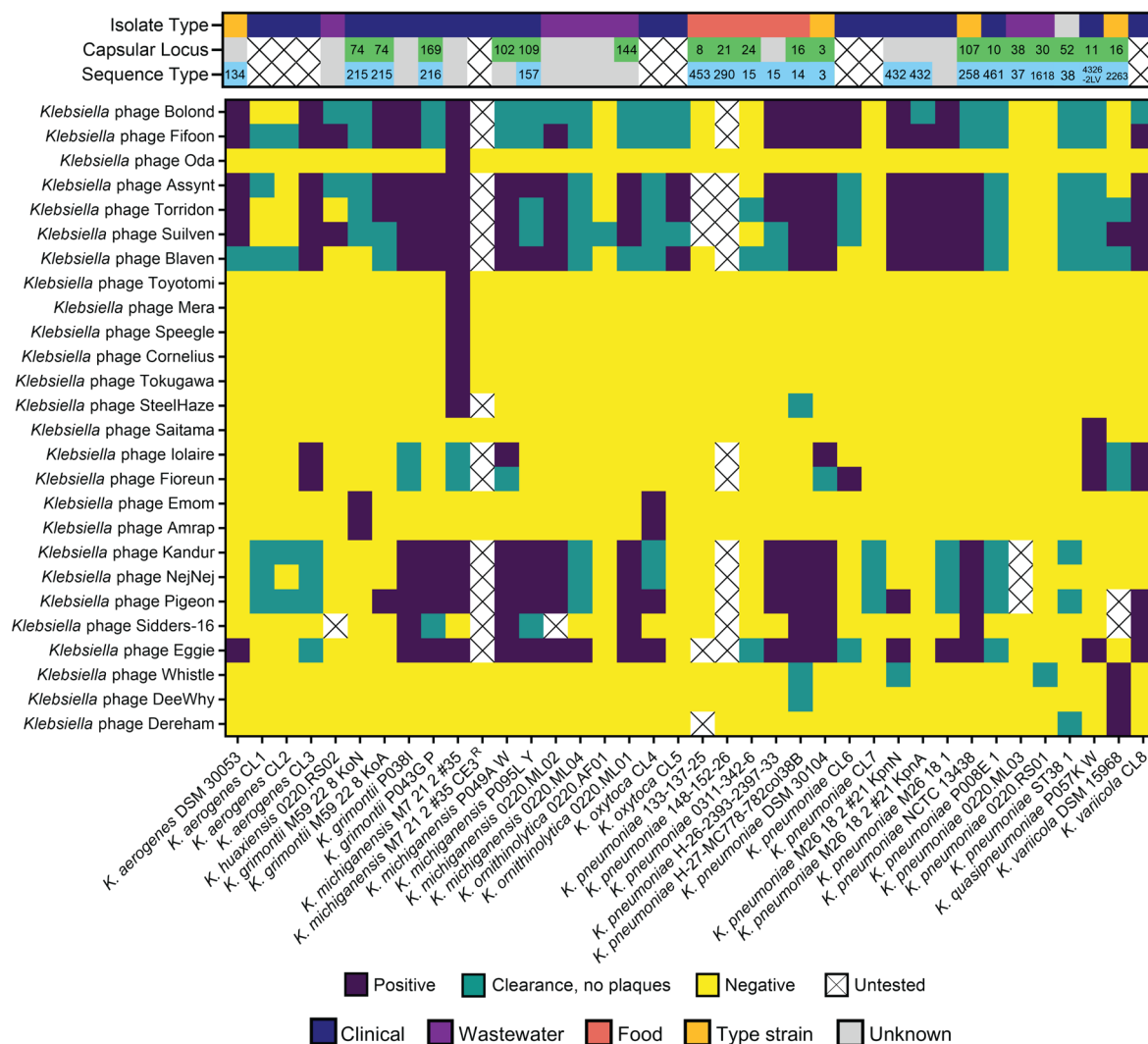
Figure 4.3. Heatmap for host range of all 26 genetically distinct virulent phages by plaque assay against a diverse range of *Klebsiella* spp. Top panel shows isolate type, capsular type, and sequence type of the *Klebsiella* host strains. The source of each bacterial isolate is given as isolate type, with grey being unknown source. Capsular loci determined by Kaptive and/or Kleborate, green; unknown, low, no match confidence, grey. Sequence type (ST) determined by MLST, blue; unknown or incomplete matches, grey. Isolates for which no sequencing data was available or were untested are indicated by a white box with a cross. Bottom panel shows host range heatmap. Productive infection (positive), purple; lysis without productive infection, green; no productive infection or clearance (negative), yellow.

Similarly to the slopekviruses, all jiaodaviruses exhibited broad host range, which appeared to demonstrate a particular pattern depending on the host and location in which they were isolated (**Figure 4.3**, **Appendix 10.3**). For example, jiaodaviruses Assynt, Torridon, Suilven, and Blaven (group 1) were capable of infecting the same *Klebsiella* strains that the other jiaodaviruses Kandur, NejNej, Pigeon, Sidders-16, and Eggie (group 2) were not. The

jiaodaviruses were isolated using two different species of *Klebsiella*: group 1 were isolated against *K. pneumoniae* KpnN, whereas group 2 were isolated against *K. grimontii* P038I. Some jiaodaviruses were capable of productively infecting fewer isolates than others. For example, phage Blaven was capable of infecting 15 out of 36 (41.6%) isolates tested, whereas Assynt was capable of infecting 19 out of 35 (54.3%) isolates. Phage Sidders-16 however, was only capable of infecting 7 out of 34 (20.6%) isolates tested, whereas Eggie could infect 18 out of 35 (51.4%) isolates. Sidders-16 was the most divergent of the jiaodaviruses in terms of host range. Interestingly, none of the jiaodaviruses were capable of productively infecting the WT *K. grimontii* KoN, and only three phages, Assynt, Torridon (both group 1), and Pigeon (group 2) were capable of productively infecting the SCV *K. grimontii* KoA. Moreover, the host range of the group 2 jiaodaviruses was variable for the WT *K. pneumoniae* KpnN, with only Eggie and Pigeon capable of productively infecting it. However, none of the group 2 jiaodaviruses were capable of infecting the SCV *K. pneumoniae* KpnA.

Another two phages Iolaire and Fioreun were isolated from the same wastewater plant and against the same host (*K. quasipneumoniae* P057K W) (**Appendix 10.3**). Both were considered broad host range, being capable of productively infecting a minimum of four different species of *Klebsiella* each (**Figure 4.3**, **Appendix 10.4**). Iolaire was capable of productively infecting five different *Klebsiella* species out of 36 isolates (13.9%), and caused lysis without productive infection against another three. Fioreun was capable of productively infecting four different *Klebsiella* species out of 36 isolates (11.1%), and caused lysis without productive infection against another five. The ability for both phages to infect multiple species of *Klebsiella* may be more useful in phage therapy than phages that can infect more isolates of same species.

Phages DeeWhy and Dereham exhibited a narrow host range, capable of productively infecting just one *Klebsiella* strain (their isolation host) out of the 38 and 37 isolates tested, respectively (**Figure 4.3**, **Appendix 10.4**). Both phages were also capable of causing lysis without productive infection against *K. pneumoniae* H-27-MC778-782col38B and *K. pneumoniae* ST38 1, respectively.

For eight *Klebsiella* strains, no phage in the current collection was capable of productive infection, including two *K. aerogenes* strains (CL1 and CL2), one *K. ornithinolytica* (0220.AF01), and five *K. pneumoniae* strains, of which one was foodborne, and the rest were clinical (**Figure 4.3**). Of the *K. pneumoniae* isolates, one was a clinical sample for which no sequencing data were available (CL7), clinical isolate *K. pneumoniae* P008E 1, wastewater isolate *K. pneumoniae* 0220.RS01, and unknown isolate *K. pneumoniae* ST38 1. No correlation was found between the antibiotic resistance status, number of phage defence mechanisms, or prophage content predicted for three of these four *K. pneumoniae* isolates (see **Chapter 3**). Indeed, both P008E 1 and ST38 1 were associated with higher

AMR predictions (identified as MDR and XDR, respectively), and had an above average number of defence systems (n = 15 and 12, respectively), whereas 0220.RS01 was neither MDR, and had below average defence systems predicted (n = 7). Moreover, 0220.RS01 only had one plasmid predicted, whereas the other two *K. pneumoniae* isolates were predicted to harbour five each. Whilst it is unclear as to why none of the 26 phages could infect the four *K. pneumoniae* strains, it is feasible that there are bacterial defence systems present that could confer protection against phage infection from those in this collection. More likely, it is probable that the phage collection is not diverse enough across the genus and unable to infect as the isolates lack the particular receptor or epitope of a receptor to facilitate infection.

Additionally, a further four *Klebsiella* isolates were not infected by any of the phages, but the data is inconclusive as some phages have not been tested against the isolates due to time constraints (**Figure 4.3**). These included the *K. michiganensis* Oda-resistant variant (*K. michiganensis* M7 Oda^R) derived from the clinical *K. michiganensis* M7 WT, two foodborne *K. pneumoniae* isolates, and the wastewater-derived *K. pneumoniae* 0220.ML03 isolate.

Together, these data show that host range can vary widely, even if the phages are genetically similar. Phages Bolond and Fifoon were initially thought to be the same phage having been isolated from the same sample (see **Chapter 6**), but against two different hosts. However, their host range suggested the two phages were distinct, perhaps encoding different depolymerases within their tail proteins, allowing them to target different hosts due to their different capsular types. Some phages can indeed encode multiple depolymerase domains within their tail fibre and/or spike structures that are capable of degrading multiple capsular types (Hsieh *et al.*, 2017; Pan *et al.*, 2017; Cai *et al.*, 2019; Latka *et al.*, 2019). However, much of the research on phage-encoded depolymerases has focused on the capsular types of *K. pneumoniae*, and it is unclear whether multiple depolymerase activity would increase the host range beyond *K. pneumoniae* capsular types. Lourenço *et al.* (2023) have shown that using acapsular *K. pneumoniae* mutants can select for phages with an even broader host range since they do not rely on the capsule for infection. Moreover, the rate of resistance emergence was slower for phages selected against acapsular mutants when compared to encapsulated *K. pneumoniae* strains (Lourenço *et al.*, 2023). Given that the capsule is the main barrier to overcome for many phages to access *Klebsiella* receptors (Beamud *et al.*, 2023), selection of phages that do not rely on it could result in more efficacious application of phages against this highly variable and adaptable bacterium (Lourenço *et al.*, 2023).

Some phages were capable of infecting WT *Klebsiella* strains, but not the SCVs, and *vice versa*.

Lysis without productive infection was observed among all the T4-like (*Straboviridae*) phages, and the unclassified siphoviruses Iolaire and Fioreun. The webervirus DeeWhy, unclassified siphovirus Dereham, and unclassified myovirus SteelHaze could only cause lysis without productive infection against one isolate each. Whistle was the only przondovirus capable of causing lysis without productive infection. Lysis-from-without is a phenomenon whereby phages adsorb to the bacterial cell in high numbers and are able to penetrate the cell wall resulting in lysis, but without the production of phage progeny (Abedon, 2011). However, lysis-from-without is often distinguished from generalised lysis without productive infection, where for the latter the bacterial cell is overwhelmed by high phage numbers, but cell death is not adsorption-induced (i.e. adsorption-independent) (Abedon, 2011). In this study, the term lysis without productive infection is used to describe generalised phage-mediated lysis irrespective of adsorption dependence. It is thus acknowledged that the molecular mechanism for the lysis without productive infection observed during host range assays may not reflect the definition in its truest sense.

Lysis-from-without is usually limited to phages with larger genomes, and has been well-characterised among the T-even phages (such as *Straboviridae* members) (Molineux, 2006). It is hypothesised that the straboviruses in this study exhibited lysis-from-without (i.e. adsorption-induced), whereas the other phages possibly induced lysis without productive infection that could be either adsorption-induced or adsorption-independent. Lysis-from-without in straboviruses can be further complicated by delayed adsorption to secondary receptors, resulting in false-negatives (Abedon, 2011). Such false-negatives could generate vastly different observations in lysis-from-without amongst the *Straboviridae* phages in the collection.

Where lysis-from-without does occur, phages are more likely to be "trained" (i.e. pre-adapted) to target the host cell and increase the host range (Friman *et al.*, 2016; Kortright *et al.*, 2019; Pirnay *et al.*, 2023). Training is achieved through co-evolution assays, which selects for phages with increased infectivity (Friman *et al.*, 2016; Kortright *et al.*, 2019). Training phages could be particularly useful within phage therapy settings. For example, high-risk or global problem clones of *Klebsiella* spp. for which no phage can productively infect could be targeted if phages within that collection have previously demonstrated lysis-from-without. Previous studies have shown positive clinical outcomes using pre-adapted phages (Friman *et al.*, 2016; Pirnay *et al.*, 2023), including in a case of MDR *K. pneumoniae* wound infection (Eskenazi *et al.*, 2022).

Overall, no link could be established between ST or capsular type and host range for any phages tested. However, many strains had non-typable ST and/or capsular types, possibly due to the draft genomes or because they were non-*K. pneumoniae* isolates. It would be interesting to see if a pattern emerged with testing of further strains, and/or determining ST and capsular types through improved genome assembly of the bacterial isolates.

4.4.2 Efficiency of plaquing

Phage Bolond was isolated against *K. michiganensis* M7 WT, but over time the plaques became more turbid and the phage would not amplify well (< 10^8 PFU/mL), suggesting this phage was not particularly efficient in this host. Phage Fifoon exhibited similar behaviour with its isolation host *K. pneumoniae* KpnN, with plaques becoming more turbid and difficult to count, but less so than Bolond. To test this further, titre and efficiency of plaquing (EOP) assays were performed for Bolond and Fifoon against several hosts informed by the host range data (**Figure 4.4A-B**). EOP assays were performed against five hosts, but the plaques for both phages against *K. michiganensis* M7 WT were too turbid to count accurately, and therefore this host was excluded.
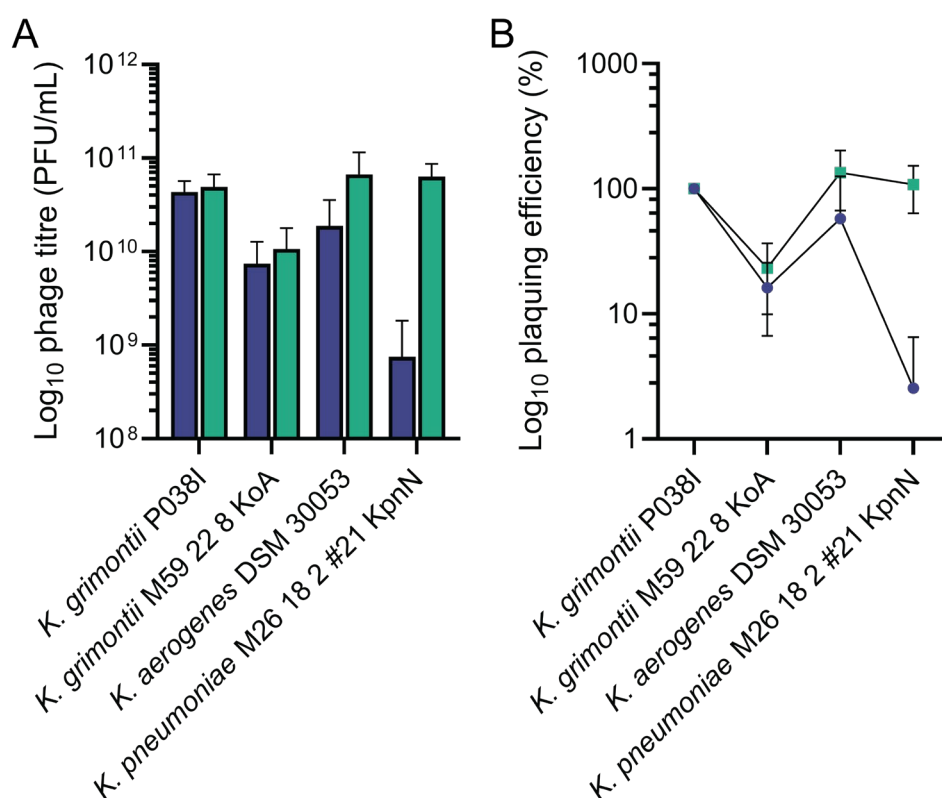
Figure 4.4. Titre and efficiency of plaquing for *Klebsiella* phages Bolond and Fifoon against four clinical *Klebsiella* isolates. (A) Titre in PFU/mL and **(B)** EOP relative to the propagation host *K. grimontii* P038I for Bolond (purple) and Fifoon (green). Data represent mean ± SD, *n* = 3 except for Fifoon tested against *K. pneumoniae*, *n* = 2. Differences between phage-host pairs PFU/mL were analysed by one-way ANOVA corrected for multiple comparisons (Tukey method) and were not statistically significant.

These data revealed that for phage Bolond, *K. grimontii* P038I was the most efficient host, with a titre of 4.33×10^{10} PFU/mL (**Figure 4.4A-B**). This phage was also efficient against *K. aerogenes*, with a titre of 1.89×10^{10} PFU/mL but demonstrated only 57.7% efficiency compared to the propagation host. The efficiency of Bolond against *K. grimontii* KoA and *K. pneumoniae* KpnN was reduced by more than 84% (7.42×10^9 PFU/mL) and 97% (7.50×10^8 PFU/mL) in comparison to *K. grimontii* P038I, respectively.

Fifoon also demonstrated high efficiency against *K. grimontii* with a titre of 4.90×10^{10} PFU/mL. However, Fifoon efficiency was highest against *K. aerogenes*, with a titre of 6.67×10^{10} PFU/mL (134.8%) (**Figure 4.4A-B**). The isolation host for Fifoon, *K. pneumoniae* KpnN was also high at 6.34×10^{10} PFU/mL (108%), but again only two replicates were included because of the difficulty in counting the plaques. The efficiency of Fifoon against *K. grimontii* KoA was the lowest at 1.06×10^{10} PFU/mL (21.7%) compared to *K. grimontii* P038I.

Based on these experiments, *K. grimontii* P038I was selected as the new propagation host for downstream analyses.

Although EOP can determine phage efficiency in various hosts, it cannot consider plaque morphology and the difficulties associated with counting very small, turbid plaques. Despite Fifoon having higher EOP against *K. aerogenes*, titres were still within the same order of magnitude and for propagation, a minimum of 10^{10} PFU/mL are required for long term storage. Thus, *K. grimontii* P038I was kept as the propagation host for both Bolond and Fifoon.

Whilst these data further support that both Bolond and Fifoon are genetically distinct phages, host tropism is a complex phenomenon, and genetically identical phages could demonstrate differences in host range and EOP, particularly if the bacterial host adapts to one phage and not the other during *in vitro* analyses.

4.4.3 Phage time-kill assays

To establish the potential of single-phage and combination therapy, phage time-kill assays were performed. Six phages were chosen to represent the variety of different taxa and host ranges and single-phage time-kill assays were performed with these six phages against three different hosts at different dosages.

For the single-phage experiments, phages from the same family generally exhibited similar killing against the same hosts, with some exceptions (**Figure 4.5A-L**). Phages Bolond and Fifoon (both in the genus *Slopekvirus*, and family *Straboviridae*) were effective at reducing *K. michiganensis* load to, or below, the limit of detection (LOD, defined as an OD_{600} of < 0.06) for multiplicities of inoculation (MOIs) within the range 100-0.1 and 1000-1, respectively (**Figure 4.5A-F**). At 5:00 hours post-inoculation (hpi) however, regrowth was observed, particularly those at higher MOIs. Initial bacterial growth with reduced bacterial load to an OD_{600} of ~0.5 occurred for both Bolond (**Figure 4.5A**) and Fifoon (**Figure 4.5B**) at MOIs or 0.1-0.01, and 1-0.1. Contrastingly, phage Oda killed the initial inoculum below the LOD, but regrowth was observed 5:00 hpi for all MOIs (**Figure 4.5C**).

A — Bolond ☆ (*K. michiganensis* M7 21 2 #35)
Legend: 100 (10^{10}), 10 (10^{9}), 1 (10^{8}), 0.1 (10^{7}), 0.01 (10^{6}), 0.001 (10^{5}), 0.0001 (10^{4}), 0.00001 (10^{3}), *K. michiganensis* M7 21 2 #35

B — Fifoon
Legend: 1000 (10^{11}), 100 (10^{10}), 10 (10^{9}), 1 (10^{8}), 0.1 (10^{7}), 0.01 (10^{6}), 0.001 (10^{5}), 0.0001 (10^{4}), *K. michiganensis* M7 21 2 #35

C — Oda ☆
Legend: 100 (10^{10}), 10 (10^{9}), 1 (10^{8}), 0.1 (10^{7}), 0.01 (10^{6}), 0.001 (10^{5}), 0.0001 (10^{4}), 0.00001 (10^{3}), *K. michiganensis* M7 21 2 #35

D — Assynt
Legend: 100 (10^{10}), 10 (10^{9}), 1 (10^{8}), 0.1 (10^{7}), 0.01 (10^{6}), 0.001 (10^{5}), 0.0001 (10^{4}), 0.00001 (10^{3}), *K. michiganensis* M7 21 2 #35

E — SteelHaze ☆
Legend: 1000 (10^{11}), 100 (10^{10}), 10 (10^{9}), 1 (10^{8}), 0.1 (10^{7}), 0.01 (10^{6}), 0.001 (10^{5}), 0.0001 (10^{4}), *K. michiganensis* M7 21 2 #35

F — Eggie
Legend: 1000 (10^{11}), 100 (10^{10}), 10 (10^{9}), 1 (10^{8}), 0.1 (10^{7}), 0.01 (10^{6}), 0.001 (10^{5}), 0.0001 (10^{4}), *K. michiganensis* M7 21 2 #35

G — Bolond
Legend: 100 (10^{10}), 10 (10^{9}), 1 (10^{8}), 0.1 (10^{7}), 0.01 (10^{6}), 0.001 (10^{5}), 0.0001 (10^{4}), 0.00001 (10^{3}), *K. grimontii* P038I

H — Fifoon
Legend: 100 (10^{10}), 10 (10^{9}), 1 (10^{8}), 0.1 (10^{7}), 0.01 (10^{6}), 0.001 (10^{5}), 0.0001 (10^{4}), 0.00001 (10^{3}), *K. grimontii* P038I

I — Eggie ☆
Legend: 100 (10^{10}), 10 (10^{9}), 1 (10^{8}), 0.1 (10^{7}), 0.01 (10^{6}), 0.001 (10^{5}), 0.0001 (10^{4}), 0.00001 (10^{3}), *K. grimontii* P038I

J — Bolond
Legend: 1 (10^{8}), 0.1 (10^{7}), 0.01 (10^{6}), 0.001 (10^{5}), 0.0001 (10^{4}), 0.00001 (10^{3}), 0.000001 (10^{2}), 0.0000001 (10^{1}), *K. pneumoniae* KpnN

K — Fifoon ☆
Legend: 10 (10^{9}), 1 (10^{8}), 0.1 (10^{7}), 0.01 (10^{6}), 0.001 (10^{5}), 0.0001 (10^{4}), 0.00001 (10^{3}), 0.000001 (10^{2}), *K. pneumoniae* KpnN

L — Assynt ☆
Legend: 100 (10^{10}), 10 (10^{9}), 1 (10^{8}), 0.1 (10^{7}), 0.01 (10^{6}), 0.001 (10^{5}), 0.0001 (10^{4}), 0.00001 (10^{3}), *K. pneumoniae* KpnN

Right margin labels: *K. michiganensis*, *K. grimontii*, *K. pneumoniae*

Axis labels: Bacterial load (OD$_{600}$); Time (h:m)

98
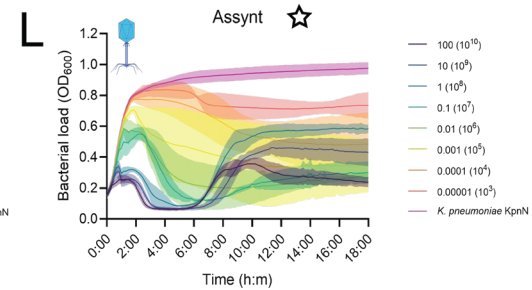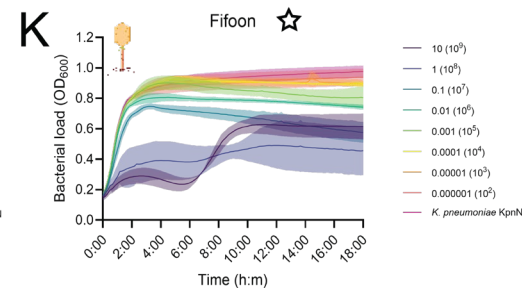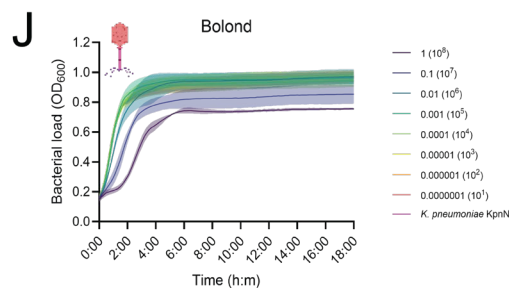
Figure 4.5. Time-kill assays for six phages against three clinical *Klebsiella* isolates at different MOIs. Time-kill assays for phages **(A)** Bolond (orange), **(B)** Fifoon (yellow), **(C)** Oda (pink), **(D)** Assynt (blue), **(E)** SteelHaze (purple), and **(F)** Eggie (green) against *K. michiganensis* M7 WT (top panel). Time-kill assays for **(G)** Bolond, **(H)** Fifoon, and **(I)** Eggie against *K. grimontii* P038I (middle panel). Killing kinetics for **(J)** Bolond, **(K)** Fifoon, and **(L)** Assynt against *K. pneumoniae* KpnN (bottom panel). Bacterial load is expressed as OD_{600}. Key represents MOIs with average PFU across all replicates. Stars denote the original isolation host. Data represent mean (solid lines) ± SD (shaded area), $n = \geq 2$.

Phages Assynt (**Figure 4.5D**) and Eggie (**Figure 4.5F**) (both in the genus *Jiaodavirus*, and family *Straboviridae*) exhibited similar killing to Bolond and Fifoon, but with more pronounced killing below the LOD from 1:30-2:00 hpi. Additionally, a slightly longer lag period before bacterial regrowth occurred at 6:00-7:00 hpi for both phages. Phage Eggie was unable to kill *K. michiganensis* M7 WT effectively at the lower MOIs (0.001-0.0001), whereas this was only true for Assynt at the lowest MOI (0.00001). Moreover, Assynt reduced bacterial load at 0.001 MOI to an OD_{600} of 0.3 at 18:00 hpi, whereas regrowth to an OD_{600} of 0.5-0.6 occurred for all other MOIs. Similarly to Bolond and Fifoon, an increased lag period was also observed for *K. michiganensis* M7 WT following inoculation with phage SteelHaze, with bacterial load at or below the LOD from 1:30-5:00 hpi for all MOIs in the range 1000-1 (**Figure 4.5E**).

Phages Bolond and Fifoon exhibited similar killing against their propagation host *K. grimontii* P038I (**Figure 4.5G-H**) as they did against *K. michiganensis* M7 WT, with the lowest MOIs having the least effect on reducing bacterial load. Whilst the lag phase for Eggie was more variable it also exhibited similar killing to Bolond and Fifoon (**Figure 4.5I**).

Having previously shown to have a lower EOP against *K. pneumoniae* KpnN, it was unsurprising that Bolond did not significantly affect bacterial load, where regrowth occurred within 2:00 hpi across all MOIs tested (**Figure 4.5J**). Contrastingly, Fifoon performed slightly better, but only at the two highest MOIs, which still resulted in regrowth after 6:00 hpi (**Figure 4.5K**). Assynt had a variable effect on bacterial load, which was highly dependent on MOI (**Figure 4.5L**). For example, MOIs 100-1 were all similar in their lag phase with variable regrowth, whereas initial growth was observed at MOIs 0.1-0.01, followed by a reduction, and then supressed regrowth.

Phages Bolond and Fifoon exhibited differences in host range and time-kill assays, but were genetically similar (see **Chapter 6**). Because an accurate genome had been assembled for Bolond but not Fifoon, Bolond was selected in combination experiments. To test the hypothesis that combinations of phages into cocktails may suppress bacterial load more than single-phage experiments, several phages were tested in two- and three-phage combinations. Phage combinations were determined based on single-phage time-kill assays, host range, and genomic diversity (see **Chapter 6**). Additionally, ratios of phage combinations were determined based on the length of their infection cycle, that is, phages

that were known to complete their infection cycles faster were diluted more relative to the other. Phages were diluted to adjust MOIs with the same starting concentration.

Phage Bolond and Oda exhibited very different killing at their respective MOIs and were combined at different ratios (**Figure 4.6A-C**). Interestingly, the lag phase for regrowth of the bacteria was extended at all ratios of Bolond to Oda at all MOIs (**Figure 4.6A-C**) compared to the single-phage treatments for both (**Figure 4.5A**). Moreover, the lower the concentration of Oda relative to Bolond (**Figure 4.6A-C**), the longer the lag phase, particularly at the higher MOIs, where regrowth occurred between 8:00 hpi and 11:00 hpi at a 10:1 ratio (**Figure 4.6C**), compared to 5:00 hpi for both alone. However, there was high variability in bacterial load in both the technical and biological replicates. Inadequate mixing prior to adding phages was a possibility and efforts were made to mix well during repeats. Additionally, samples of Bolond and Oda at a ratio of 1:1 were sent for TEM imaging (**Appendix 10.5**) to see if any other virion particle interaction could be causing such variability. TEM revealed no viable Bolond particles, where all those imaged had empty capsids (**Appendix 10.5**). Whilst there was some aggregation of Bolond, this is usually a factor of storing high-titre phages. Phage Oda could not be distinguished from any of the debris in the images (**Appendix 10.5**).

Combinations of SteelHaze (unclassified myovirus) and either Assynt or Eggie (both jiaodaviruses) revealed that both combinations were able to decrease bacterial load and supress regrowth (**Figure 4.7A-B**). SteelHaze in combination with Assynt increased the lag phase by several hours when compared to either phage alone (**Figure 4.7A**). Moreover, at 0.1-0.01 MOI, bacterial regrowth did not exceed an OD_{600} of 0.2 from 2:00-18:00 hpi (**Figure 4.7A**). SteelHaze in combination with Eggie exhibited similar killing, particularly with considerable reduction and suppression of growth from 3:00-18:00 hpi at 0.1 MOI (**Figure 4.7B**).

Two three-phage combinations included adding either Bolond or Oda to both SteelHaze and Eggie (**Figure 4.7C-D**). The highest MOIs 100 and 10 reduced bacterial load, but resulted in some regrowth after 8:00-10:00 hpi for both sets of combinations. At the lowest MOIs, the three-phage combination containing Bolond had less of an effect on bacterial load than the other MOIs over the course of the experiment (**Figure 4.7C**). Bacterial load was reduced somewhat at MOIs 0.01 and 0.001 at 18:00 hpi. This three-phage combination was the most efficacious at MOIs of 1-0.1, where the bacterial lag phase was extended over the entire 18 h experiment, where growth was suppressed to an OD_{600} of ≈2.
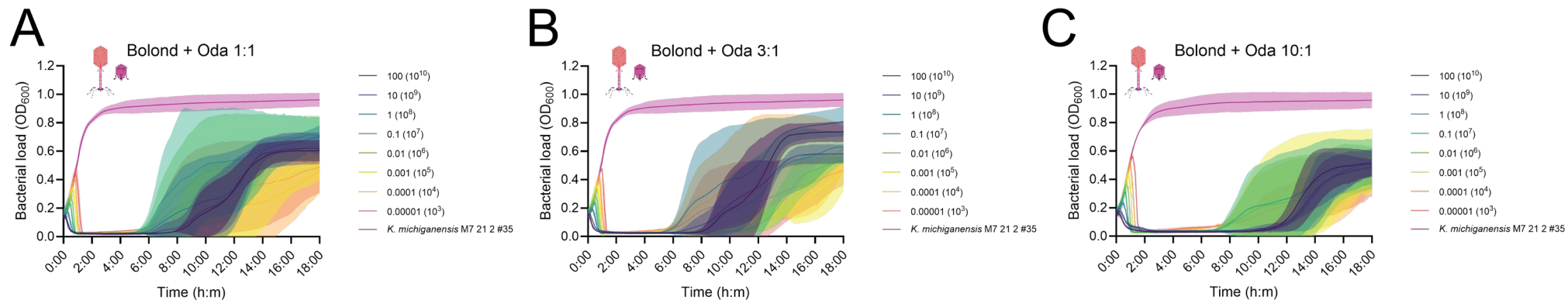
Figure 4.6. Killing kinetics of phages Bolond and Oda in combination at three different ratios against a clinical *K. michiganensis* M7 21 3 #35 WT isolate at different MOIs. Phages Bolond (orange) and Oda (pink) in combination at a **(A)** 1:1 ratio a **(B)** 3:1 ratio and a **(C)** 10:1 ratio. Phage Oda starting MOI was diluted relative to phage Bolond phage following phage enumeration experiments (*n* = 3). Bacterial load is expressed as OD_{600}. Key represents MOIs with average PFU across all replicates. Data represent mean (solid lines) ± SD (shaded area), *n* = ≥ 2.

The three-phase combination containing Oda instead of Bolond overall showed a greater reduction in bacterial load across all MOIs (**Figure 4.7D**). Some regrowth was observed at the highest MOIs of 100 and 10 (phage Oda 10 and 1). Interestingly, the bacterial lag phase was extended over the entire 18-hour experiment, where growth was suppressed to an OD$_{600}$ of ~1-1.5 across all the remaining MOIs.

Together, these data suggested that two-phase combinations were more effective at killing and suppressing bacterial growth over a wider range of MOIs than each individual phage alone. Moreover, three-phase combinations were even more efficacious than both single-phase and two-phase combinations. More importantly however, was the type of phages used in combination therapy. That is, the three-phase combinations most efficacious were SteelHaze, Eggie, and Oda, which represented completely different families of phages. The other three-phase combination contained two phages from the same family (*Straboviridae*) but different genera. This combination was not as efficacious, possibly because the two phages from the same family competed with each other to bind to the phage receptors.

Combination therapy can increase host range, particularly if the combination targets a diverse set of strain (Gibson *et al.*, 2019; Romero-Calle *et al.*, 2019; Molina *et al.*, 2022). Moreover, combination therapy can reduce the emergence of bacterial resistance, which is generally achieved through selecting phages that target distinct host receptors to one another (Gibson *et al.*, 2019; Romero-Calle *et al.*, 2019; Molina *et al.*, 2022). Bacterial resistance to phages can manifest in many different ways, but one of the most common is through receptor loss or downregulation (Hsieh *et al.*, 2017; Tan *et al.*, 2019; Hesse *et al.*, 2020; Maffei *et al.*, 2021; Pires *et al.*, 2021). The accumulation of multiple mutants to resist phage infections not only occurs at reduced rates during combination therapy, but the fitness costs may result in concomitant reduction in virulence or re-sensitisation to antibiotics (Cai *et al.*, 2018; Gibson *et al.*, 2019; Hesse *et al.*, 2020). It is unclear whether bacterial regrowth during time-kill assays were caused by stochastic resistance, or other factors such as persister cells. Persister cells are metabolically inactive subpopulations that occur in almost all bacterial communities, and can resist both antibiotic and phage therapy since they rely on functional cellular machinery to exert their effects (Maffei *et al.*, 2024).

High variability in bacterial load in both the technical and biological replicates when phages Bolond and Oda were combined into a two-phase cocktail could be explained by random (stochastic) events, leading to bacterial resistance at different time frames and with different effects on fitness (Wright *et al.*, 2019). Whilst no particle interaction was observed in the TEM, some phages can interact with one another, resulting in antagonistic effects *in vitro* that reduce stability (Hsu *et al.*, 2019; Kortright *et al.*, 2019; Hesse *et al.*, 2020).
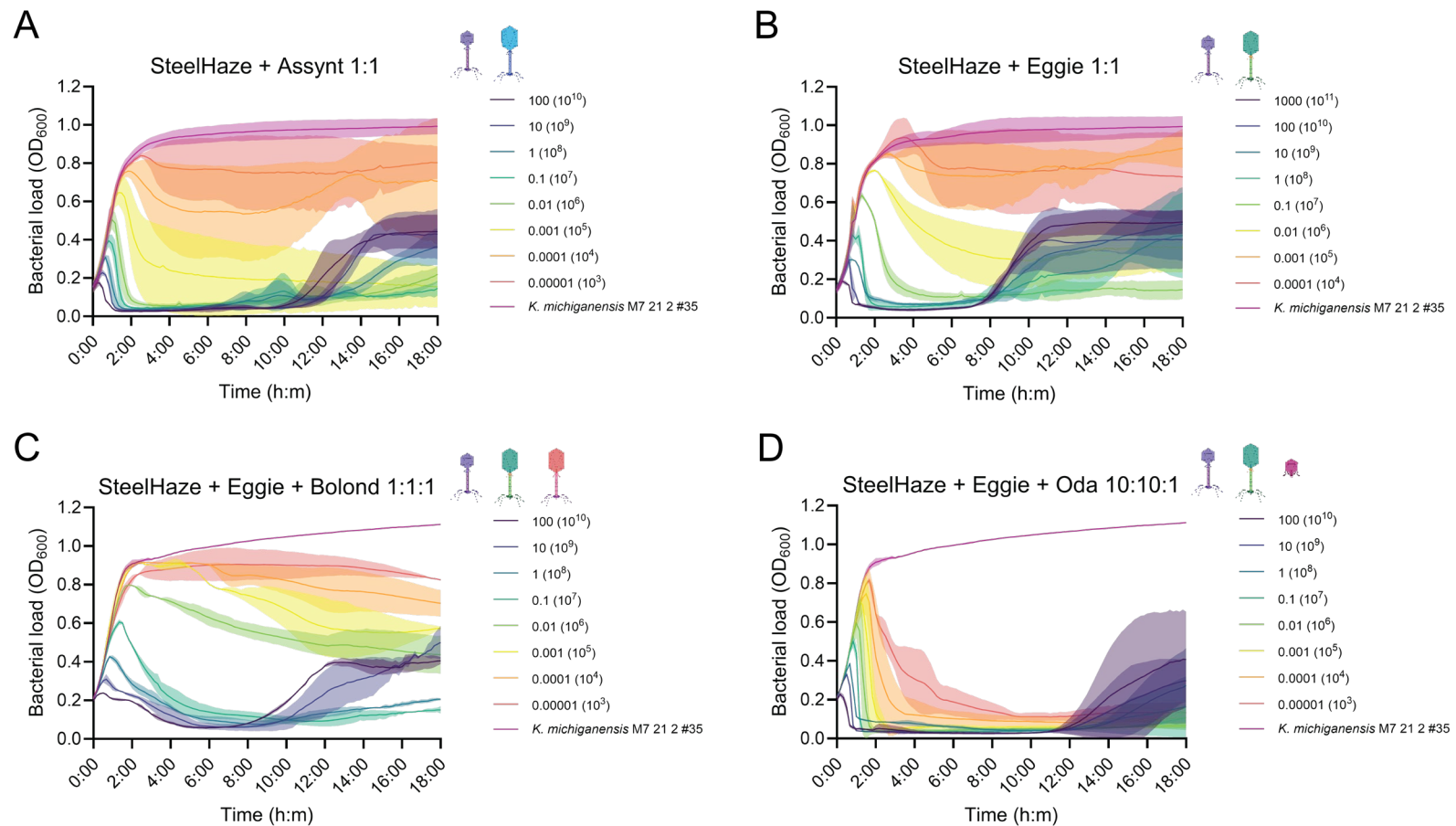
Figure 4.7. Time-kill assays for two- and three-phage combinations against a clinical *K, michiganensis* M7 21 2 #35 WT isolate at different MOIs.
(A) Phages SteelHaze (purple) and Assynt (blue) in combination at a 1:1 ratio. **(B)** Phages SteelHaze (purple) and Eggie (green) in combination at a 1:1 ratio.
(C) Phages SteelHaze (purple), Eggie (green) and Bolond (orange) in combination at a 1:1 ratio. **(D)** Phages SteelHaze (purple), Eggie (green) and Oda (pink) in combination at a 10:10:1 ratio. Phage Oda starting MOI was diluted relative to phage Bolond phage following phage enumeration experiments (*n* = 3). Bacterial load is expressed as optical density OD_{600}. Key represents MOIs with average PFU across all replicates. Data represent mean (solid lines) ± SD (shaded area), *n* = ≥ 2.

4.5 Phage-antibiotic interactions

To investigate phage-antibiotic interactions, first antibiotic susceptibility was established by minimum inhibitory concentration (MIC) testing on two different strains of the same isolate, *K. michiganensis* M7 WT and the phage-resistant mutant, *K. michiganensis* M7 Oda^R. The MIC results then informed the phage-antibiotic synergy assays.

4.5.1 Antibiotic susceptibility testing

Antibiotic susceptibility testing by MIC is defined as the minimum concentration of a compound that inhibits visible growth of the bacterial strain following overnight (18:00 hours) incubation. MICs were performed with the *K. michiganensis* M7 WT and M7 Oda^R to determine if either antibiotic or phage resistance resulted in any cross-resistance between the two.

A total of 12 antibiotics spanning eight different classes (β-lactams included penicillins, cephalosporins, carbapenems, and their combinations with tazobactam) were tested against *K. michiganensis* M7 WT (**Figure 4.8A**). These included carbenicillin (CAR), piperacillin (PIP) (both penicillins in the β-lactam class), piperacillin/tazobactam (TZP), chloramphenicol (CHL, a phenicol), ciprofloxacin (CIP, a quinolone), colistin (CST, a polymyxin), cefotaxime (CTX, a cephalosporin in the β-lactam class), kanamycin (KAN, an aminoglycoside), meropenem (MEM, a carbapenem in the β-lactam class), streptomycin (STR, an aminoglycoside), tetracycline (TET, a tetracycline), and trimethoprim (TMP).

The bacterium was susceptible to all antibiotics except for the penicillins without tazobactam, as expected since it encoded *bla*$_{OXY}$, which is responsible for conferring resistance to the penicillin class of antibiotics. The WT strain showed slight differences in MICs against two different penicillins, CAR and PIP, but was resistant to both. Structural differences between CAR and PIP could have caused the differences in the MICs for both penicillins. The addition of tazobactam to PIP (TZP) resulted in a four-fold reduction in MIC however, suggesting that the WT strain did not contain determinants capable of resisting β-lactamase inhibitors, such as carbapenemases (see **Table 1.2**). MICs for *K. michiganensis* M7 Oda^R showed differences in the MICs of CST, MEM, and TET (**Figure 4.8B**) to the WT, none of which resulted in changes in antibiotic susceptibility. *K. michiganensis* M7 Oda^R was also resistant to CAR and PIP, but not TZP.

These data showed no significant variance in the MICs for *K. michiganensis* M7 Oda^R when compared to the WT, suggesting that phage resistance did not result in cross-resistance to the panel of antibiotics tested, except for STR, which was not tested against *K. michiganensis* M7 Oda^R.
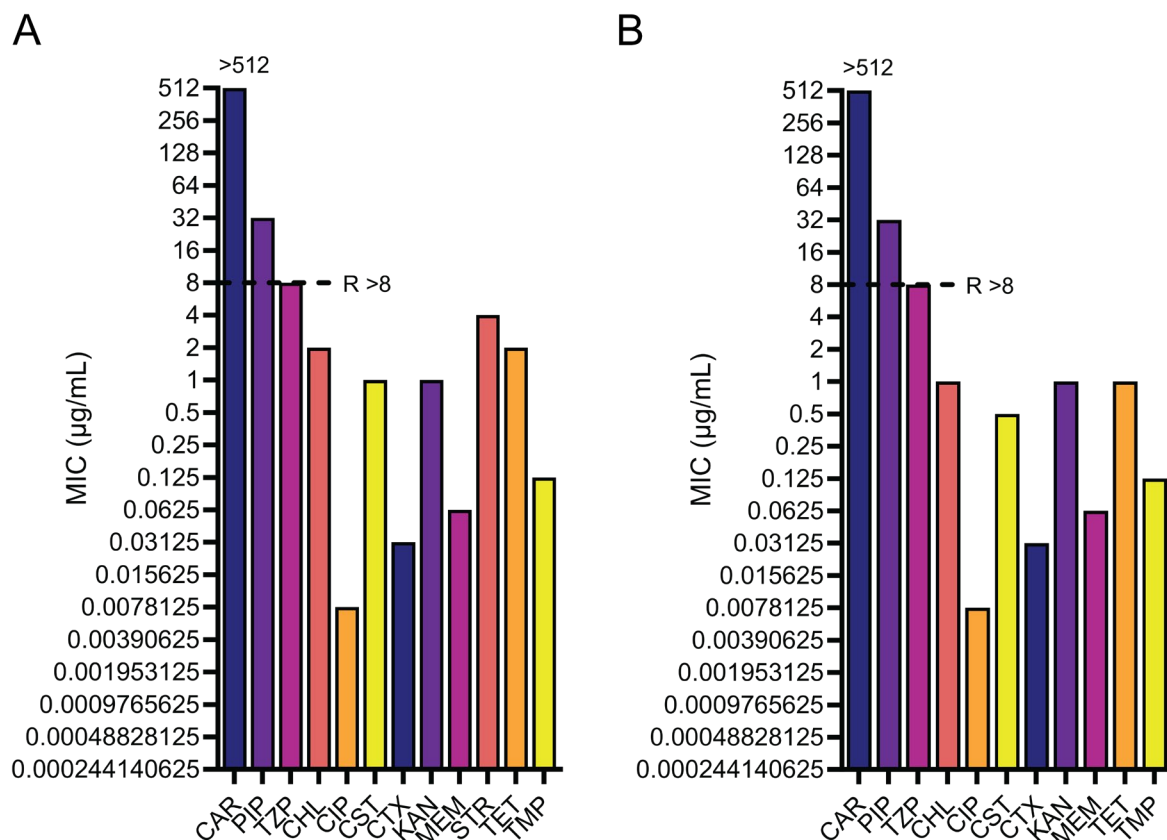
Figure 4.8. Antibiotic susceptibility testing for a range of clinically relevant antibiotics against the WT and a phage resistant strain of *K. michiganensis* M7 21 2 #35. (A) MICs for 12 antibiotics tested against *K. michiganensis* M7 WT, spanning eight different classes. **(B)** MICs for 11 different antibiotics tested against *K. michiganensis* M7 Oda^R, spanning seven different classes. R, resistance as determined by EUCAST breakpoints (EUCAST, 2023). Data represent mean ± SD, *n* = 3.

4.5.2 Phage-antibiotic synergy testing

Phage-antibiotic interactions can be measured by a range of methods. Here, synergy testing was performed using checkerboard assays because of their simplicity in that they were performed very similarly to MIC assays (**Figure 4.9A-B**). All were performed with *K. michiganensis* M7 WT as the host.

Antibiotic type (bactericidal versus bacteriostatic) and class can exhibit different interactions when combined with different phages. To test this hypothesis, synergy testing was performed with four phages spanning four different genera; Oda, Bolond, SteelHaze, and Eggie; and six different antibiotics representative of four classes, with CHL as the only bacteriostatic compound against a single bacterial isolate. The fractional inhibitory concentration (FIC) index is used to define phage-antibiotic interactions and is expressed as a range of 0.5 to 4 (ESCMID, 2000; Wind *et al.*, 2015). Phage-antibiotic interactions were quantified by calculating the FIC indices from the MIC and minimum inhibitory MOI (MIM) determined by the checkerboard assays (summarised in **Table 4.3**).

No effect was observed for phage Oda in combination with the bactericidal antibiotic, CIP (**Figure 4.9C**). As expected, Oda in combination with the bacteriostatic antibiotic CHL had an antagonistic effect (**Figure 4.9D**). However, it was possible that the starting MOI for Oda was too high, masking any potential synergistic or additive effects, particularly with bactericidal antibiotics.

A Schematic representation of an MIC assay. **B** Schematic representation of a checkerboard assay.

C CHL (μg/mL)

Oda (MOI)	4	2	1	0.5	0.25	0.125	0.0625	0.03125	0.015625	0.007813	0.003906	0
10	0.001	0.000	0.003	-0.002	0.002	-0.003	0.002	0.000	0.001	0.003	0.000	0.000
1	0.001	0.000	0.001	0.000	0.000	-0.003	0.002	-0.004	0.002	0.002	-0.003	0.000
0.1	0.003	0.002	0.001	0.000	0.000	0.002	0.001	-0.001	0.057	-0.001	0.001	0.007
0.01	-0.001	0.003	0.000	-0.003	0.004	0.000	0.000	0.001	0.001	0.001	-0.001	0.001
0.001	0.002	0.001	0.002	0.015	0.000	-0.001	0.003	0.001	0.026	0.001	-0.001	0.002
0.0001	0.001	0.003	0.001	0.066	0.073	0.024	-0.001	0.266	0.007	0.004	0.002	0.004
0.00001	0.001	0.007	0.247	0.000	0.134	0.000	0.390	0.005	0.004	0.140	0.394	0.004
0	0.002	0.011	0.558	0.887	1.096	1.178	1.241	1.237	1.257	1.285	1.195	1.237

D CIP (μg/mL)

Oda (MOI)	0.03125	0.015625	0.007813	0.003906	0.001953	0.000977	0.000488	0.000244	0.000122	0.00006	0.000031	0
10	0.001	-0.002	-0.002	0.010	0.000	0.003	-0.002	-0.001	-0.001	0.001	-0.005	0.002
1	0.003	0.011	0.004	-0.003	0.005	-0.001	-0.002	-0.002	0.003	0.001	-0.002	-0.003
0.1	-0.001	-0.003	0.002	0.003	0.001	0.000	-0.001	0.004	0.000	0.002	-0.002	0.001
0.01	0.002	-0.001	0.004	0.003	0.004	0.002	0.005	0.002	0.005	0.001	-0.001	0.002
0.001	0.002	0.005	-0.001	0.005	0.002	0.004	0.065	0.004	0.005	0.001	0.273	-0.001
0.0001	0.005	-0.001	0.002	0.019	0.010	0.002	0.188	0.003	0.002	0.001	0.380	0.000
0.00001	0.006	0.003	-0.001	0.005	0.072	0.506	0.341	0.415	0.010	0.003	0.524	0.003
0	0.007	0.006	0.674	0.975	0.977	0.881	0.908	0.991	1.007	1.088	1.014	0.996

Figure 4.9. Phage-antibiotic synergy testing with phage Oda. (A) Schematic representation of an MIC assay. **(B)** Schematic representation of a checkerboard assay, which is similar to an MIC assay. Pink lines denote no antibiotic control (right column) and no phage control (bottom row). 0 denotes the no antibiotic, no phage control. Black lines denote where the MIC_A and MIM_B alone were determined. Black well outline denotes were the MIC and MIM in combination were read (MIC_{AC} and MIM_{BC}, respectively). Growth (OD_{600}) of *K. michiganensis* M7 21 2 #35 WT in MHB supplemented with doubling dilutions of **(C)** CHL and **(D)** CIP, in combination with ten-fold dilutions of phage Oda. Growth was determined as an $OD_{600} \geq 0.06$, indicated by the black outline and is correlated to colour intensity. Data represent mean from $n = 2$. Computer-readable data available in **Appendix 10.6**.

A — CAR (µg/mL)

Bolond (MOI)	1024	512	256	128	64	32	16	8	4	2	1	0
10	-0.016	0.256	0.202	0.122	0.364	0.304	0.298	0.268	0.693	0.502	0.740	0.657
1	-0.002	0.219	0.314	0.253	0.347	0.651	0.657	0.332	0.397	0.501	0.359	0.481
0.1	-0.005	0.367	0.336	0.580	0.252	0.148	0.260	0.127	0.144	0.394	0.458	0.368
0.01	0.053	0.337	0.414	0.530	0.450	0.206	0.113	0.223	0.525	0.307	0.687	0.437
0.001	0.058	0.346	0.434	0.996	0.686	0.764	0.696	0.609	0.622	0.643	0.950	0.854
0.0001	-0.018	0.268	0.406	1.033	0.917	1.185	1.140	1.012	1.178	0.974	0.915	0.958
0.00001	-0.109	0.270	0.514	1.138	1.163	1.144	1.129	0.972	1.195	1.019	0.959	1.106
0	-0.014	0.203	0.414	1.133	1.169	1.242	1.159	1.199	1.211	1.164	1.123	1.202

B — CTX (µg/mL)

Bolond (MOI)	0.125	0.0625	0.03125	0.015625	0.007813	0.003906	0.001953	0.000977	0.000488	0.000244	0.000122	0
10	-0.003	-0.004	0.082	0.359	0.162	0.091	0.260	0.068	0.122	0.317	0.392	0.329
1	0.002	-0.004	0.160	0.243	0.179	0.178	0.275	0.330	0.127	0.369	0.188	0.125
0.1	0.012	0.022	0.227	0.260	0.237	0.117	0.072	0.195	0.266	0.115	0.231	0.323
0.01	0.002	0.023	0.472	0.495	0.437	0.181	0.188	0.022	0.092	0.241	0.153	0.163
0.001	0.001	0.040	0.365	0.396	0.543	0.751	0.473	0.428	0.553	0.500	0.399	0.525
0.0001	0.000	0.093	0.095	0.302	0.767	1.070	0.795	0.606	0.704	0.808	0.926	0.977
0.00001	0.010	0.135	0.362	0.277	0.769	1.093	1.106	1.028	1.134	1.025	1.068	1.167
0	0.001	-0.001	0.001	0.232	0.636	1.144	0.998	1.178	1.180	1.180	1.191	1.199

C — CHL (µg/mL)

Bolond (MOI)	4	2	1	0.5	0.25	0.125	0.0625	0.03125	0.015625	0.007813	0.003906	0
10	0.001	0.002	0.125	0.600	0.293	0.207	0.040	0.091	0.476	0.350	0.492	0.532
1	0.002	0.003	0.032	0.667	0.047	0.015	0.078	0.283	0.036	0.318	0.199	0.023
0.1	0.000	0.002	0.071	0.262	0.205	0.176	0.289	0.200	0.015	0.166	0.374	0.083
0.01	0.002	-0.001	0.308	0.342	0.444	0.098	0.057	0.165	0.365	0.445	0.253	0.412
0.001	0.000	0.002	0.417	0.711	0.217	0.131	0.152	0.246	0.278	0.371	0.327	0.414
0.0001	0.002	0.005	0.510	0.458	0.497	0.307	0.146	0.068	0.084	0.099	0.157	0.118
0.00001	0.003	0.000	0.424	0.926	0.890	0.515	0.348	0.489	0.604	0.784	0.692	0.629
0	0.004	0.011	0.209	0.482	1.191	0.605	1.271	1.222	1.268	1.272	1.283	1.206

D — CIP (µg/mL)

Bolond (MOI)	0.03125	0.015625	0.007813	0.003906	0.001953	0.000977	0.000488	0.000244	0.000122	0.00006	0.000031	0
10	-0.013	-0.015	0.187	0.177	0.302	0.452	0.494	0.251	0.240	0.462	0.157	0.497
1	-0.011	0.180	0.405	0.255	0.297	0.283	0.363	0.540	0.351	0.491	0.287	0.335
0.1	-0.009	0.137	0.581	0.475	0.256	0.379	0.407	0.263	0.407	0.482	0.239	0.042
0.01	-0.012	0.124	0.708	0.654	0.262	0.309	0.282	0.191	0.419	0.211	0.519	0.339
0.001	-0.010	0.143	0.803	0.985	0.970	0.488	0.419	0.673	0.587	0.374	0.748	0.763
0.0001	-0.009	0.121	0.806	0.788	0.742	1.161	0.978	1.067	1.228	1.164	1.072	1.024
0.00001	-0.015	0.072	0.818	1.041	1.129	1.101	0.865	1.237	1.255	1.187	1.103	1.200
0	-0.011	-0.011	0.713	0.961	1.015	1.037	1.173	1.109	1.204	1.195	1.176	1.212

Figure 4.10. Phage-antibiotic synergy testing with phage Bolond. Growth (OD_{600}) of *K. michiganensis* M7 21 2 #35 WT in MHB supplemented with doubling dilutions of **(A)** CAR, **(B)** CTX, **(C)** CHL, and **(D)** CIP, in combination with ten-fold dilutions of phage Bolond. Growth was determined as an $OD_{600} \geq 0.06$, indicated by the black outline and is correlated to colour intensity. Data represent mean from *n* = 2. Computer-readable data available in **Appendix 10.7**.

No effect was observed for phage Bolond in combination with any antibiotic (**Figure 4.10A-D**), possibly due to the nature of the kinetics for this particular phage against its original isolation host *K. michiganensis* M7 WT. Testing a higher MOI (> 10), could potentially result in observations of effects, but providing a such large number of phages in a clinical setting would make it economically unfeasible.

Increased growth inhibition was observed for phage SteelHaze in four out of the six antibiotics tested (**Figure 4.11A-F**). SteelHaze had an additive effect in combination with both CAR and CIP (**Figure 4.11A**, **4.11D**), and no effect in combination with KAN (**Figure 4.11E**). Interestingly, SteelHaze had a synergistic effect in combination with both the bactericidal antibiotic CTX (**Figure 4.11B**) and the bacteriostatic antibiotic CHL (**Figure 4.11C**), reducing the MIC by eight-fold and four-fold, respectively. No effect could be determined for SteelHaze in combination with MEM (**Figure 4.11F**).

A CAR (µg/mL)

SteelHaze (MOI)	1024	512	256	128	64	32	16	8	4	2	1	0
10	0.003	-0.002	0.001	0.001	0.003	0.003	0.001	0.002	0.003	0.004	-0.013	0.005
1	0.004	0.002	0.006	0.001	0.007	0.003	0.003	0.006	0.003	0.005	0.006	0.002
0.1	0.005	0.002	0.006	0.003	0.004	0.003	0.001	0.005	0.004	0.002	-0.007	0.001
0.01	0.001	0.005	0.007	0.011	0.013	0.328	0.005	0.010	0.231	0.283	0.016	0.075
0.001	0.002	0.009	0.056	0.044	0.235	0.210	0.157	0.021	0.327	0.453	0.036	0.031
0.0001	0.058	0.220	0.050	0.213	0.340	0.107	0.055	0.247	0.574	0.240	0.130	0.379
0.00001	0.004	0.294	0.174	0.050	1.221	0.937	0.931	1.014	1.346	0.518	1.380	1.012
0	-0.004	0.355	0.321	0.669	1.240	0.983	1.447	1.425	1.429	1.455	1.439	0.984

B CTX (µg/mL)

SteelHaze (MOI)	0.125	0.0625	0.03125	0.015625	0.007813	0.003906	0.001953	0.000977	0.000488	0.000244	0.000122	0
10	0.000	-0.004	0.000	0.000	0.000	-0.005	0.000	-0.001	-0.002	-0.004	-0.004	0.002
1	0.002	0.002	-0.001	0.002	0.000	0.023	-0.001	0.000	0.020	-0.001	-0.002	-0.001
0.1	0.002	0.004	-0.001	0.000	0.001	0.000	0.014	0.002	0.000	0.001	0.002	0.003
0.01	0.003	0.001	-0.001	0.004	0.003	0.008	0.112	0.122	0.015	0.005	0.233	0.119
0.001	0.003	0.014	0.010	0.013	0.021	0.191	0.199	0.046	0.124	0.515	0.451	0.013
0.0001	0.003	0.002	0.069	0.184	0.043	0.056	0.158	0.343	0.081	0.236	0.073	0.315
0.00001	0.001	0.001	0.047	0.254	0.607	1.250	0.076	0.946	0.527	0.721	0.920	0.686
0	0.000	0.005	0.378	0.401	0.755	1.352	1.431	1.391	1.467	1.477	1.461	1.226

C CHL (µg/mL)

SteelHaze (MOI)	4	2	1	0.5	0.25	0.125	0.0625	0.03125	0.015625	0.007813	0.003906	0
10	0.000	0.000	0.000	0.003	0.001	0.004	0.000	-0.006	0.003	0.040	0.021	-0.003
1	0.001	0.000	0.007	0.001	0.001	0.002	0.003	0.004	0.003	-0.002	-0.002	0.005
0.1	0.000	-0.034	0.002	0.001	0.001	0.003	0.017	0.000	0.435	0.002	0.517	0.007
0.01	0.002	-0.001	-0.001	0.004	0.064	0.248	0.319	0.000	0.305	0.238	0.375	0.343
0.001	-0.002	0.006	-0.001	0.005	0.287	0.095	0.587	0.481	0.354	0.488	0.479	0.183
0.0001	0.001	0.008	-0.003	0.018	0.535	0.299	0.532	0.944	0.282	0.571	0.626	0.135
0.00001	-0.002	0.007	0.574	0.458	1.060	1.159	1.149	1.190	1.215	0.752	1.238	0.669
0	0.001	0.007	0.274	0.902	1.205	1.309	1.276	1.295	1.304	1.316	1.283	1.210

D CIP (µg/mL)

SteelHaze (MOI)	0.03125	0.015625	0.007813	0.003906	0.001953	0.000977	0.000488	0.000244	0.000122	0.00006	0.000031	0
10	0.002	0.009	0.003	0.002	-0.003	0.001	-0.002	0.002	0.002	0.000	-0.001	-0.002
1	0.000	0.000	0.002	0.000	-0.002	0.001	-0.002	-0.001	-0.001	0.000	-0.001	0.001
0.1	0.001	-0.004	-0.003	-0.004	0.000	-0.003	-0.001	-0.002	0.001	0.022	0.000	0.011
0.01	-0.002	-0.003	-0.002	-0.002	0.000	0.007	0.003	0.004	0.005	0.051	0.000	0.002
0.001	-0.002	-0.002	-0.001	0.026	0.021	0.017	0.098	0.026	0.078	0.215	0.017	0.232
0.0001	0.001	0.027	0.000	0.053	0.360	0.228	0.227	0.109	0.189	0.036	0.165	0.422
0.00001	-0.002	-0.002	-0.002	0.394	0.712	0.140	0.895	1.205	0.999	0.841	0.882	1.364
0	-0.002	-0.001	-0.001	0.782	1.198	1.271	1.325	1.264	1.397	1.436	1.450	0.826

E KAN (µg/mL)

SteelHaze (MOI)	4	2	1	0.5	0.25	0.125	0.0625	0.03125	0.015625	0.00781	0.003906	0
10	0.002	0.001	0.002	0.005	0.002	0.001	0.002	0.003	0.004	0.005	0.006	0.004
1	0.000	-0.003	0.002	0.000	-0.001	0.001	0.002	0.001	0.002	0.003	0.000	0.000
0.1	0.000	0.032	0.009	0.000	0.032	0.010	0.043	0.011	0.003	0.008	0.000	0.005
0.01	0.008	0.064	0.338	0.374	0.144	0.142	0.066	0.211	0.130	0.003	0.011	0.505
0.001	-0.005	0.075	0.700	0.529	0.308	0.187	0.245	0.187	0.457	0.027	0.363	0.234
0.0001	-0.002	0.112	0.164	0.324	0.405	0.058	0.504	0.584	0.149	0.366	0.578	0.267
0.00001	0.003	0.994	0.963	0.334	1.242	0.842	1.155	1.024	0.611	0.360	0.744	0.693
0	0.000	0.923	1.176	1.210	1.235	1.138	1.138	1.163	1.132	1.142	1.150	1.178

F MEM (µg/mL)

SteelHaze (MOI)	0.5	0.25	0.125	0.0625	0.03125	0.015625	0.007813	0.003906	0.001953	0.00098	0.000488	0
10	-0.047	-0.024	-0.020	0.503	-0.023	0.541	-0.024	0.001	0.462	-0.023	-0.024	-0.020
1	-0.024	0.416	0.454	0.009	-0.024	-0.018	-0.026	0.260	0.002	0.268	-0.032	-0.022
0.1	-0.024	0.310	-0.025	-0.024	0.160	-0.021	-0.022	-0.021	-0.013	-0.020	0.714	-0.015
0.01	-0.028	-0.022	-0.024	0.738	-0.025	0.013	0.197	-0.022	-0.013	-0.018	-0.022	0.439
0.001	0.481	0.460	0.895	0.502	-0.013	0.301	-0.008	-0.007	0.339	-0.003	0.530	0.503
0.0001	0.065	0.442	0.460	0.604	0.172	0.134	0.536	0.676	0.438	0.057	0.731	0.801
0.00001	0.189	0.595	0.642	1.017	0.877	1.129	0.915	1.175	1.183	0.232	1.040	1.170
0	0.056	1.169	1.215	1.200	1.124	1.165	1.192	1.157	1.188	1.192	1.186	1.201

Figure 4.11. Phage-antibiotic synergy testing with phage SteelHaze. Growth (OD_{600}) of *K. michiganensis* M7 21 2 #35 WT in MHB supplemented with doubling dilutions of **(A)** CAR, **(B)** CTX, **(C)** CHL, **(D)** CIP, **(E)** KAN, and **(F)** MEM, in combination with ten-fold dilutions of phage SteelHaze. Growth was determined as an $OD_{600} \geq 0.06$, indicated by the black outline and is correlated to colour intensity. Data represent mean from *n* = 4. Computer-readable data available in **Appendix 10.8**.

Figure 4.12. Phage antibiotic synergy testing with phage Eggie. Growth (OD_{600}) of *K. michiganensis* M7 21 2 #35 WT in MHB supplemented with doubling dilutions of **(A)** CAR, **(B)** CTX, **(C)** CHL, and **(D)** CIP, in combination with ten-fold dilutions of phage Eggie. Growth was determined as an $OD_{600} \geq 0.06$, indicated by the black outline and is correlated to colour intensity. Data represent mean from *n* = 2. Computer-readable data available in **Appendix 10.9**.

A — CAR (µg/mL)

Eggie (MOI)	1024	512	256	128	64	32	16	8	4	2	1	0
10	-0.011	-0.008	-0.011	-0.010	-0.006	-0.008	-0.009	-0.012	-0.010	-0.009	-0.010	-0.006
1	-0.010	-0.012	-0.009	-0.012	-0.011	-0.009	-0.011	-0.009	-0.008	-0.018	-0.009	-0.009
0.1	-0.009	-0.008	0.037	-0.010	-0.007	-0.010	-0.009	-0.008	-0.008	-0.008	-0.007	-0.007
0.01	-0.009	-0.008	-0.007	-0.005	-0.007	-0.006	-0.010	-0.005	-0.003	0.006	-0.005	-0.007
0.001	0.009	0.003	0.003	0.005	0.017	0.009	0.004	0.004	0.002	0.002	-0.002	-0.003
0.0001	-0.009	0.005	0.069	0.200	0.595	0.434	0.309	0.021	0.525	0.546	0.487	0.554
0.00001	0.003	0.035	0.309	0.771	1.078	0.723	0.679	1.190	0.699	1.021	0.641	0.907
0	0.022	0.152	0.289	0.897	1.197	1.190	1.207	0.621	1.205	1.208	1.189	1.219

B — CTX (µg/mL)

Eggie (MOI)	0.125	0.0625	0.03125	0.015625	0.007813	0.003906	0.001953	0.000977	0.000488	0.000244	0.000122	0
10	-0.004	0.000	-0.003	-0.004	-0.002	-0.003	-0.004	-0.003	0.000	-0.002	0.001	-0.004
1	-0.002	0.001	-0.002	-0.003	-0.003	-0.001	-0.003	0.000	-0.006	0.000	-0.002	0.002
0.1	0.004	0.002	0.006	0.002	0.001	0.001	0.001	0.004	0.001	0.000	0.005	0.000
0.01	0.005	-0.004	0.001	0.001	0.001	0.006	0.004	0.000	-0.001	0.006	0.003	0.003
0.001	0.003	0.000	0.043	0.016	0.016	0.018	0.041	0.017	0.009	0.008	0.016	0.006
0.0001	-0.001	0.002	0.013	0.055	0.039	0.136	0.051	0.201	0.301	0.028	0.012	0.043
0.00001	-0.007	0.001	0.004	0.126	0.287	0.924	0.947	1.121	1.133	0.921	0.818	0.416
0	0.003	0.009	-0.002	0.206	0.805	1.195	1.250	1.218	1.198	1.174	1.112	1.073

C — CHL (µg/mL)

Eggie (MOI)	4	2	1	0.5	0.25	0.125	0.0625	0.03125	0.015625	0.007813	0.003906	0
10	-0.002	-0.002	-0.001	-0.003	-0.002	0.000	0.004	0.005	0.002	0.000	-0.001	0.001
1	0.001	0.000	0.000	-0.004	-0.003	0.001	0.001	0.001	-0.001	0.000	-0.001	0.003
0.1	0.001	0.004	0.089	0.004	-0.002	0.000	0.000	0.006	0.002	-0.005	-0.003	0.000
0.01	-0.001	0.007	0.304	0.173	0.006	0.008	0.003	0.006	0.018	0.001	0.005	0.003
0.001	0.001	0.007	0.454	0.260	0.304	0.012	0.117	0.283	0.442	0.008	0.649	0.007
0.0001	0.016	0.012	0.590	0.682	0.122	0.251	0.019	0.070	0.007	0.198	0.564	0.091
0.00001	0.000	0.008	0.545	0.708	0.591	0.983	0.959	0.836	1.046	0.617	0.959	0.620
0	-0.001	0.012	0.639	0.967	1.038	1.175	1.223	1.221	1.242	1.257	1.231	1.186

D — CIP (µg/mL)

Eggie (MOI)	0.03125	0.015625	0.007813	0.003906	0.001953	0.000977	0.000488	0.000244	0.000122	0.00006	0.000031	0
10	-0.005	-0.001	-0.005	-0.006	-0.005	0.033	-0.007	-0.008	-0.006	-0.004	-0.003	-0.003
1	-0.004	-0.004	-0.006	-0.005	-0.006	-0.006	-0.006	-0.004	0.033	-0.006	-0.013	-0.004
0.1	-0.006	-0.005	-0.005	0.020	-0.003	0.030	0.010	-0.004	-0.004	-0.003	0.007	-0.014
0.01	-0.006	-0.005	-0.004	0.000	0.021	0.030	0.039	-0.004	0.000	0.021	-0.002	-0.003
0.001	-0.005	-0.007	0.023	0.229	0.089	0.328	0.013	0.093	0.063	0.177	0.384	0.000
0.0001	-0.004	0.009	0.260	0.322	0.147	0.230	0.317	0.118	0.580	0.279	0.277	0.375
0.00001	-0.007	-0.005	0.748	0.710	0.609	0.867	1.080	0.619	0.784	1.094	0.936	1.064
0	-0.004	-0.004	0.730	0.981	1.095	1.070	1.136	1.200	1.191	1.229	1.260	1.198

Increased growth inhibition was observed for phage Eggie in combination with CAR only, having an additive effect (**Figure 4.12A**). Eggie in combination with CHL had an antagonistic effect (**Figure 4.12C**), whereas no effect was observed when tested in combination with CTX (**Figure 4.12B**) and CIP (**Figure 4.12D**).

The combinatorial effects of two or more antimicrobial agents can be quantified by calculating FIC indices that indicate whether the combination is synergistic, additive, no effect, or antagonistic (ESCMID, 2000; Wind *et al.*, 2015). Each phage-antibiotic combination exhibited differences in their ability to inhibit bacterial growth. The bacteriostatic antibiotic CHL had no effect in combination with Oda, a synergistic effect with SteelHaze, and an antagonistic effect with Eggie. Bacteriostatic antibiotics have been shown to be more antagonistic when combined with phages due to their mechanism of action. That is, bacteriostatic antibiotics limit bacterial growth without killing the cell (Maier *et al.*, 2021), where in CHL the mechanism of action is through inhibition of protein synthesis

(Malmir *et al.*, 2018). This mechanism of action often has a detrimental effect on the phage infection cycle, since many phages require the very cellular machinery inhibited by the antibiotic to produce phage progeny (Gu Liu *et al.*, 2020). Phage Eggie in the family *Straboviridae* was hypothesised to redirect the host polymerases to produce phage progeny, which could explain the highly antagonistic effect of CHL (Hinton, 2010). This was not the case for Oda, which encodes its own RNA polymerase (see **Chapter 6**). However, the rapidity with which killing occurred with phage Oda alone likely skewed the antibiotic effects: the MOI needed to be adjusted to a top dilution of 1 or even 0.1 to see any full effect with CHL and CIP.

Table 4.3. FIC indices for clinically relevant antibiotics in combination with four phages against *K. michiganensis* M7 21 2 #35 WT. MIM values were treated as if for doubling dilutions for calculating the FIC indices. No FIC indices could be calculated for Bolond.

Combination	FIC Index	Effect
Klebsiella phage Oda		
Oda + CHL	2.5	Antagonistic
Oda + CIP	1.25	No effect
Klebsiella phage SteelHaze		
SteelHaze + CAR	0.5625	Additive
SteelHaze + CTX	0.375	Synergistic
SteelHaze + CHL	0.375	Synergistic
SteelHaze + CIP	0.625	Additive
SteelHaze + KAN	1.5	No effect
SteelHaze + MEM	nd	nd
Klebsiella phage Eggie		
Eggie + CAR	0.75	Additive
Eggie + CTX	1.25	No effect
Eggie + CHL	8.5	Antagonistic
Eggie + CIP	1.5	No effect

nd, not determined

Both CTX and CHL augmented the activity of phage SteelHaze in a dose-dependent manner. Synergy most commonly occurs between antibiotics with a similar mechanism of action (Bollenbach, 2015). However, when combined with phages, phage-antibiotic synergy is usually the result of two different mechanisms of action. These data suggest that synergy can occur between both bactericidal (CTX) and bacteriostatic (CHL) antibiotics and phages, and likely depend on the phage-antibiotic pair. Similarly, CAR had an additive effect when combined with both SteelHaze and Eggie, which was dose-dependent.

CIP is a rapidly-acting bactericidal antibiotic, but had antagonistic and no effects when combined with different phages. KAN has a similar mechanism of action to CHL (both protein synthesis inhibitors), where the former is bactericidal rather than bacteriostatic. However, KAN had no effect when combined with SteelHaze, whereas CHL was synergistic.

What made the checkerboards particularly challenging was that high MOIs tended to select for resistance and regrowth, whereas for antibiotics subinhibitory concentrations drove resistance selection. Thus, calculating FIC indices was challenging where the phage-only control would have high growth at the highest MOIs, meaning that direct comparisons was nearly impossible for some phages. For example, phage Bolond when tested in combination with CIP, exhibited higher killing at MOIs to 0.1, but the phage-only control exhibited growth at MOIs 10 and 1, but not 0.1. In this case calculations for phage alone (MIM_B) and in combination (MIM_{BC}) were skewed. As there were only two biological replicates (and a single technical replicate as two plates of antibiotic and phage are made at the same time), it was difficult to determine whether there are outliers in the data due to poor mixing or other interactions.

Another challenge was the use of ten-fold dilutions of phage, which made the calculations difficult. FIC indices are calculated using doubling dilutions, but using doubling dilutions of phage would have limited the range of MOIs tested, which are usually ten-fold. Thus, for ease, calculations were presumed to be doubling for both phage and antibiotics.

Phage-antibiotic combinations have been found to be highly varied and multifactorial (Gu Liu *et al.*, 2020). Importantly, the efficacy of phage-antibiotic combinations have been found to be highly dependent on not only the antibiotic mechanism of action, but on the phage infection dynamics (Gu Liu *et al.*, 2020). For example, synergism between β-lactam antibiotics and phages has been described previously, where the antibiotic mechanism of action (cell wall inhibition) potentiates phage replication driven by bacterial cell filamentation (Comeau *et al.*, 2007; Gu Liu *et al.*, 2020; Bulssico *et al.*, 2023). Other mechanisms of phage-antibiotic synergy include the use of efflux pump proteins as phage receptors, which can result in re-sensitisation of bacteria to the antibiotic in question (Chan *et al.*, 2016). Contrastingly, phages can encode proteins that are targeted by the antibiotic in addition to the target host protein that results in an antagonistic effect (Gu Liu *et al.*, 2020). The mechanism of action of bacteriostatic antibiotics is to prevent growth without killing cells (Maier *et al.*, 2021). Some bacteriostatic antibiotics achieve this by inhibiting the replicative machinery, for example, which directly affects phage replication (Pons *et al.*, 2023). Not only does this prevent virulent phages from completing their life cycle and lysing the cell, it also promotes the emergence of resistance against some phages via CRISPR-Cas immunity (Dimitriu *et al.*, 2022; Pons *et al.*, 2023).

4.6 Conclusions

A total of 26 genetically distinct phages were isolated and characterised. Many of the phages preliminarily assigned to the same family exhibited similar host range capabilities, which informed which phages to take forward for phage infection dynamics assays and phage-antibiotic synergy assays. Initially, the slopekviruses Bolond and Fifoon were

hypothesised to be the same phage, but host range analyses and growth curves against some *Klebsiella* isolates suggested they were genetically distinct phages, which has so far been corroborated by the genomic data (see **Chapter 6**). The growth curves of slopekviruses Bolond and Fifoon were similar against *K. michiganensis* WT for example, but different against *K. pneumoniae* KpnN. Similarly, the two jiaodaviruses Assynt and Eggie exhibited similar growth curves to one another against *K. michiganensis* WT. Based on the likelihood of both slopekviruses and both jiaodaviruses targeting the same receptors as one another in their respective pairs, combination growth curves were performed with just one of each (Bolond and Eggie). Phage Oda exhibited a considerably different growth curve to all the other phages tested. Being in the family *Autographiviridae*, Oda likely has a very short lifecycle, being able to complete it in less than 20 minutes. Oda also likely targeted distinct host receptors to other phages in the collection, and performed particularly well in three-phage combinations with two genetically distinct phages, SteelHaze and Eggie.

Some phage-antibiotic combinations were synergistic, whereas others were additive, had no effect, or were antagonistic. Oda in particular showed a more detrimental effect in combination with antibiotics, whereas SteelHaze was more efficacious. However, none of the two- and three-phage combinations were tested in combination with antibiotics, and multiple issues with the checkerboard assays could suggest spurious results.

Together, these data suggest that two- and three-phage combinations were more efficacious at killing and supressing regrowth of bacteria than single-phage combinations. However, such combinations together with phage-antibiotic synergy, are highly variable. Thus, phage cocktail design needs to consider multiple factors, including dosage (MOI of the phages and differential dosage of the individual phages); the infection dynamics of the phage against the targeted host(s), including life cycle, receptors, and resistance development; and the type of antibiotic if assessing phage-antibiotic synergy.

Integral to the successful application of phages therapeutically is the mechanistic understanding of phage biology within that setting, i.e. identification of the phage receptors. This information can inform clinical predictions based on the type of phage resistance that might develop and the phage co-evolution and/or fitness cost of host-mediated phage resistance.

Future work would include assessing the virulence and/or AMR status of single phage-resistant *Klebsiella* mutants following infection, and whether double-, or triple-phage resistance emerges.

CHAPTER 5: IMPLEMENTATION AND OPTIMISATION OF PHAGE AMPLIFICATION AND PURIFICATION PROTOCOLS

5.1 Introduction

For application of phages to treat recalcitrant bacterial infections, phages must be propagated (i.e. amplified) and purified to remove bacterial debris, particularly endotoxin (Sacher and Zheng, 2021; Rebula *et al.*, 2023). The amplification and purification are indeed a bottleneck in the application of phages as therapeutics, and there exists no standardised methodology currently available (Larsen *et al.*, 2023; Suleman *et al.*, 2024). Regulatory approval of any therapeutic in the UK requires production to Good Manufacturing Practice (GMP), which ensures that medicinal products are quality controlled to a particular set of internationally recognised standards (Jones *et al.*, 2023; Suleman *et al.*, 2024). The use of phages in compassionate cases i.e. those for whom all other treatment options have failed in the UK has previously relied on sourcing phages from the US and/or Belgium that have produced non-GMP high-purity, high-titre phage preparations (Dedrick *et al.*, 2019; Jones *et al.*, 2023; Suleman *et al.*, 2024).

Phage amplification can be performed on solid media by double agar overlay, or in liquid culture. However, liquid amplification of phages is essential for phage therapy, which requires scalability and allows amplification processes to be standardised more easily. Generally, liquid amplification is performed by spiking a bacterial culture with phage, followed by removal of bacterial cells and large debris by centrifugation and filtration. There are a number of different protocols for purification of phage lysates, but most begin with initial filtration followed by concentration of phage particles with, but not limited to, polyethylene glycol (PEG), followed by ultrafiltration using caesium chloride (CsCl) density gradient, tangential flow filtration (TFF), and anion-exchange chromatography, with some protocols using a combination of such methods (Smrekar *et al.*, 2011; Adriaenssens *et al.*, 2012; Bonilla *et al.*, 2016; Luong *et al.*, 2020; Larsen *et al.*, 2023; Roshankhah *et al.*, 2023).

Anion-exchange chromatography is a type of anion-exchange that purifies negatively-charged compounds by binding them to a positively charged matrix within a column (Rebula *et al.*, 2023; Roshankhah *et al.*, 2023). Unbound compounds are collected in flow through, the column is washed to elute further impurities while the protein or phage of interest remains bound until it is eluted (Smrekar *et al.*, 2011; Roshankhah *et al.*, 2023). Elution relies on the salt concentration in the elution buffer, whose negatively charged ions compete with the protein or phage for attachment to the matrix (Smrekar *et al.*, 2011; Roshankhah *et al.*, 2023). The concentration of elution buffer is increased in relation to loading buffer to elute the protein(s) of interest (Smrekar *et al.*, 2011; Vandenheuvel *et al.*, 2018). Anion-exchange combined with fast protein liquid chromatography (FPLC) offers a rapid and scalable method of purifying phages (Vandenheuvel *et al.*, 2018).

This chapter will examine the development, implementation and optimisation of phage amplification and purification protocols, with results and discussion combined for each section.

5.1.1 Aims

1. Develop and implement a pilot-scale phage amplification protocol
2. Implement and optimise a pilot-scale phage purification protocol using anion-exchange with FPLC
3. Assess endotoxin concentration of purified phage stocks

5.2 Pilot-scale amplification of phage stocks

Following phage isolation and single-plaque purification, phages were initially amplified on solid media. To minimise resources and improve scalability, phages were also amplified in liquid media in up to 50 mL bacterial culture, referred to as small-scale amplification. Amplification of phages in volumes of up to 500 mL were referred to as classical pilot-scale amplification. To develop a more scalable and standardised approach to larger-scale amplification however, a candidate phage was propagated using the Cellexus CellMaker bioreactor at a volume of 4 L. *Klebsiella* phage Fifoon was selected as the first candidate.

5.2.1 Pilot-scale amplification of phage Fifoon using a bioreactor

Based on the growth curve data for phage Fifoon against its propagation host *K. grimontii* P038I, 4 L of bacterial culture was spiked with phage at 0.001 MOI and grown for four hours to produce new phage particles. Single-use bioreactor bags within a self-contained unit allows for minimal cross-contamination of different phages to be amplified. As a side note, the bioreactor capacity allows for 8 L phage lysate to be produced in a single run, allowing amplification to be doubled without changing the conditions.

A systematic comparison between solid media amplification, small-scale liquid amplification, classical pilot-scale liquid amplification, and amplification with the bioreactor of phage Fifoon was not performed. However, a high titre was achieved following amplification with the bioreactor at $\approx 2 \times 10^{11}$ PFU/mL and was almost an order of magnitude higher to previous small-scale liquid amplification of $\approx 5 \times 10^{10}$ PFU/mL, resulting in a 300-fold increase in absolute numbers of phage particles after amplification in the bioreactor ($\sim 8 \times 10^{14}$ PFU).

The large volume of lysate produced using the bioreactor still had to be processed to remove the bacterial cells and larger debris, which is typically performed by centrifugation and filtration to produce a crude lysate that can be stored long-term, as previously described. Scalability of this step can be achieved through centrifugation of larger volumes of lysate and/or direct filtration via one or more methods (discussed below). A TFF device

was considered for filtration and concentration, but since our system was small and not automated, it would require individuals to sit for more than six hours to filter and concentrate approximately 250 mL of phage lysate, and was deemed unsuitable for this application. Thus, standard centrifugation and filtration (0.22 µm pore size) of the phage lysate was selected, and attempts were made to upscale the volumes centrifuged. However, delays in sourcing appropriate consumables and plasticware was hindered by the COVID-19 pandemic and/or Brexit, and centrifugation of lysate was performed in 50 mL batches. The capsule of *Klebsiella* spp. means that the bacteria are not particularly adherent during centrifugation, and required longer and higher spin. Additionally, the presence of the capsule also slowed filtration, since lower adherence and bacterial carryover following centrifugation would clog the filter more often, requiring regular change of the filters. Together, the poor scalability of the centrifugation and filtration of the lysate prior to purification was a considerable bottleneck following bioreactor amplification of Fifoon.

Phage Fifoon was initially selected as the first candidate for downstream analyses because it had a broader host range than Bolond, and was going to be used in phage combination assays. However, assembling a complete and accurate genome for Fifoon was difficult (see **Chapter 6**), and Bolond was used in combination assays instead. Another two or three phages were to be amplified using the protocol described here, but further bioreactor amplification was not performed due to time constraints. Thus, amplification for other phages was performed by classical pilot-scale amplification of up to 500 mL.

5.2.2 Classical pilot-scale amplification of phages

Classical pilot-scale amplification was performed for phages Bolond, Oda, Assynt, SteelHaze, and Eggie as larger volumes were needed for the phage time-kill and phage-antibiotic synergy assays (see **Chapter 4**).

Classical pilot-scale amplification of phage Assynt was performed to produce titres of at least 10^{10} PFU/mL. Whilst no direct comparison between solid media, small-scale, and classical pilot-scale amplification was performed, two batches of phage Assynt were produced by classical pilot-scale amplification, with titres measuring 1×10^{10} and 3.8×10^{11} PFU/mL, respectively. The two batches were thus taken forward for purification by anion-exchange with FPLC.

5.3 Anion-exchange with fast protein liquid chromatography for the purification of phage stocks

To develop scalable phage lysate purification system, an anion-exchange chromatography-based protocol on an FPLC system was tested with a single T4-like phage, Assynt (in the family *Straboviridae*).

Two batches of Assynt were produced to test on the high affinity (quaternary amine, QA) and low affinity (diethylamine, DEAE) columns. Both lysates were diluted 1:4 in binding buffer (20 mM Tris-HCl, pH 7.5) to reduce ionic strength and phages were eluted by step-wise gradient in elution buffer (20 mM Tris-HCl, pH 7.5, 2 M NaCl) (total volume 1 L).

Initial binding, wash and elution conditions were selected based on information from literature (Adriaenssens *et al.*, 2012), with a step-wise gradient of 5% (DEAE column) or 10% (QA column) to wash, 40% to elute, and 100% of elution buffer to remove all remaining bound particles. The process of loading phage particles onto the QA column from filtered lysate was monitored using UV and conductivity sensors (**Figure 5.1A**). Initial UV absorbance and conductivity readings while loading reflected the richness of the growth medium and the presence of phage particles, cellular debris and proteins in the lysate (**Figure 5.1A**). There were two eluting peaks observed during the 10% and 40% elution buffer steps, which equated to an ionic strength of ~0.2 M and ~0.8 M NaCl, respectively. The 10% elution peak was collected in fractions P3 A2 and P3 A3; and the 40% elution peak was collected in fractions P3 B3, P3 B4, and P3 B5 (**Figure 5.1A**). All fractions were collected per 6 mL of flow through or elution and were analysed by plaque assay, which revealed infective particles were eluted not just within the 40% peak as expected, but in all fractions, except the flow through and the column wash (data not shown) (**Figure 5.1B, Appendix 10.10**). The lack of plaquing in the flow through suggested that the column was not saturated with phages and that the initial binding conditions selected were appropriate.

The highest titres of phage in all eluted fractions were within the 10% (P3 A2 and P3 A3) and 40% fractions (P3 B3, P3 B4, and P3 B5) (**Figure 5.1B**, **Appendix 10.10**). The total phage titre that was loaded onto the column was 2.5×10^{12} in 250 mL prior to dilution (1×10^{10} PFU/mL), whereas the total number of infective particles recovered in the 10% and 40% fractions were just 12.6% (3.1×10^{11}) and 9% (2.3×10^{11}), respectively. The total number of infective particles recovered in all fractions tested was 5.6×10^{11} in a total of 120 mL, indicating the total number of infective particles recovered in all fractions was just 22.6% of the total input. Such a low recovery rate suggested significant loss of infective particles that could not be accounted for. Indeed, the remaining sum of infective particles in all remaining fractions (excluding the 10% and 40% fractions) amounted to 0.956% of total input. Moreover, despite the two UV absorbance peaks suggesting protein/particle elution, a large number of infective particles were recovered in the tail-ends of both peaks. In particular, the P3 C3 fraction showed almost double the number of infective particles compared to the previous fraction (**Figure 5.1B**, **Appendix 10.10**).
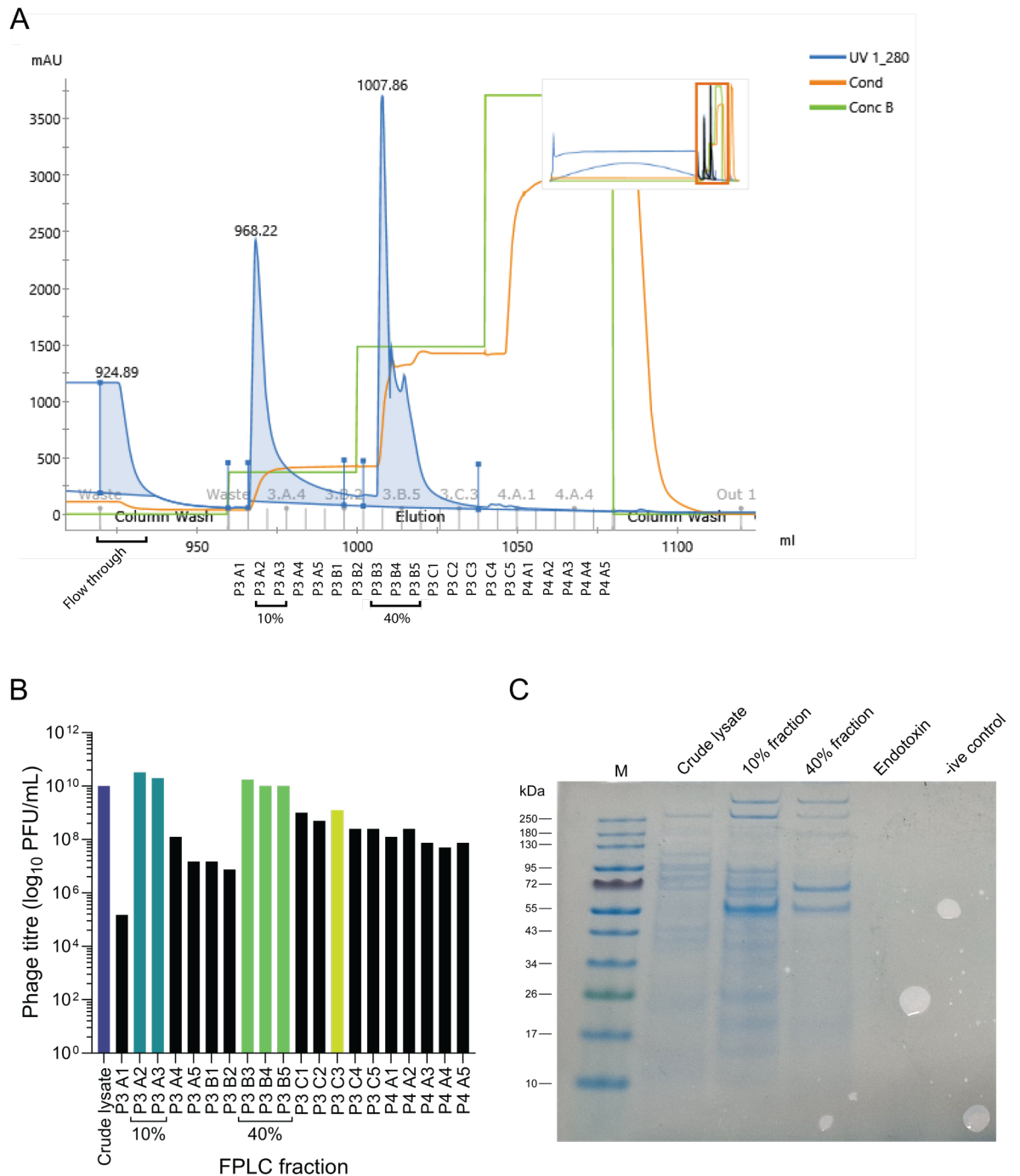
Figure 5.1. Concentration and purification of phage Assynt by anion-exchange with FPLC. (A) Chromatogram of phage purification with a CIMmultus® QA monolithic high affinity column using a step-wise gradient of elution buffer. The UV absorbance (blue line) indicated protein elution. The orange line represents conductivity. The green line represents the concentration of elution buffer with the following steps: 0%, 10%, 40%, 100% elution buffer. **(B)** Phage titres of Assynt crude lysate (purple); 10% elution peak was collected in fractions P3 A2 and P3 A3 (dark green); 40% elution peak was collected in fractions P3 B3, P3 B4, and P3 B5 (light green); the tail end of the 40% peak was collected in fraction C3 P3 (yellow-green), with remaining fractions given in black. Flow through measurements were not included. **(C)** SDS-PAGE of Assynt crude lysate (diluted 1:3), pooled Assynt FPLC 10% fraction and pooled FPLC 40% fraction. Endotoxin was used as an additional negative control, *E. coli* endotoxin O111:B4 (20 EU/mL); -ive, negative control; M, marker.

To assess the purity of the phage preparations by proxy of their protein content, an SDS-PAGE analysis was performed with the crude lysate and the 10% and 40% fractions of phages from the first run with the QA column (**Figure 5.1C**). Clear differences in the protein content of both fractions were observed, particularly with a thicker band at ~60 kDa in the 10% fraction than the 40% fraction, and thicker band at ~65 kDa in the 40% fraction (**Figure 5.1C**). Whilst there was a faint ~65 kDa band in the crude lysate, there was almost no visible ~60 kDa band in the crude lysate. What was most noticeable however, was that the 40% fraction appears to have fewer proteins than the 10% fraction, suggesting that the 40% fraction did contain fewer impurities.
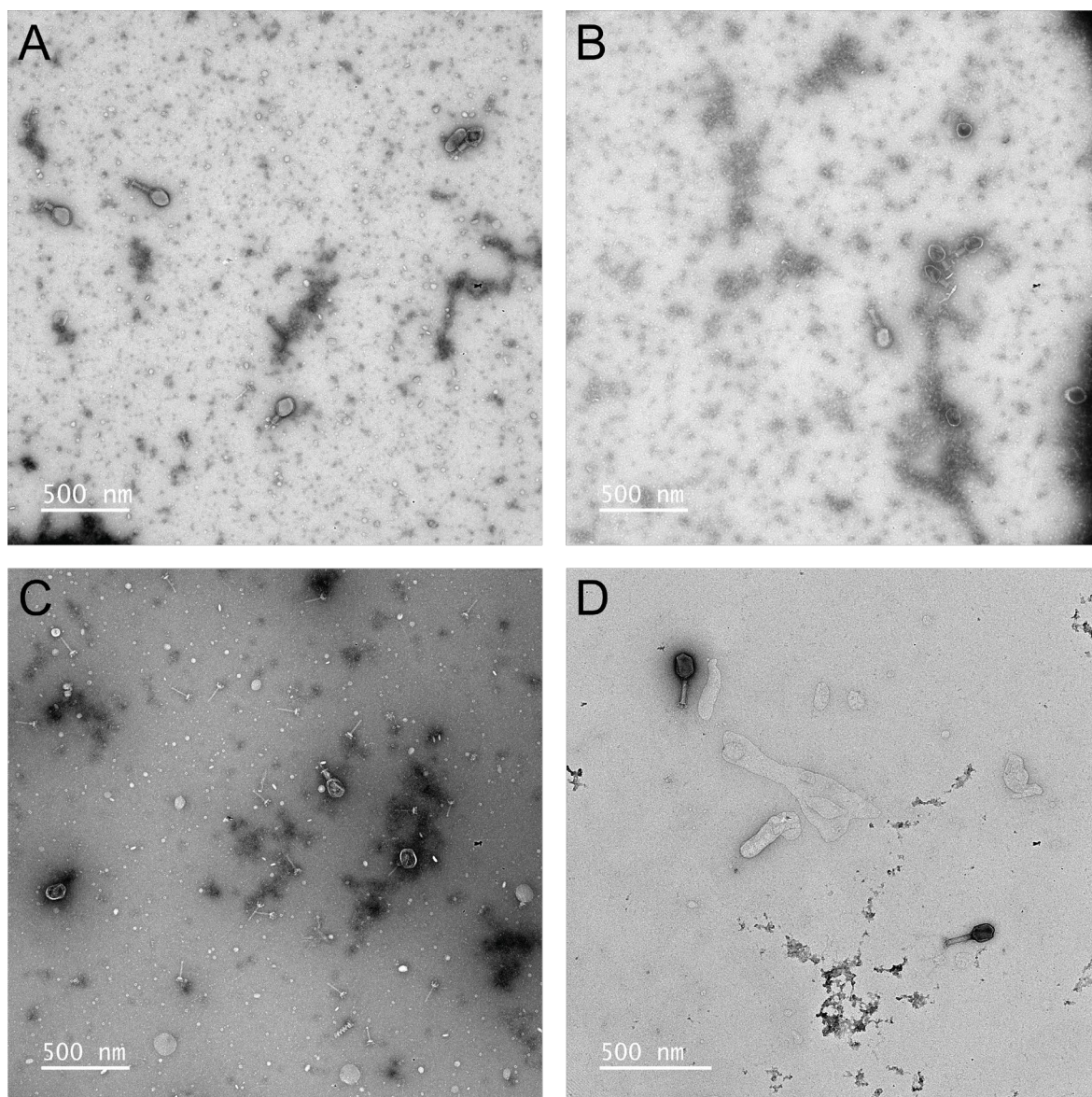
Figure 5.2. TEM micrographs of phage Assynt prior to and following anion-exchange chromatography and FPLC. TEM micrographs of **(A)** diluted (1:4) phage Assynt crude lysate; **(B)** undiluted pooled 10% FPLC fractions (P3 A2 and P3 A3); **(C)** diluted (1:4) pooled 40% FPLC fractions (P3 B3, P3 B4, and P3 B5); and **(D)** diluted (1:4) 40% peak tail fraction (P3 C3). Scale bar represents 500 nm. All micrographs were taken at 17.5 kX magnification except P3 C3 (D), which was 22.0 kX.

To further assess the phage viability of the crude lysate and FPLC fractions, TEM was performed on the crude lysate prior to anion exchange and FPLC (**Figure 5.2A**), the 10% fraction (**Figure 5.2B**), the 40% fraction (**Figure 5.2C**), and the P3 C3 fraction (**Figure 5.2D**). Interestingly, viable phage particles were observed in all micrographs, except for the 40% fraction. Viable phage particles were determined as those that had the head-tail intact, with tail fibres attached, had not started the DNA injection process (as indicated by contraction of the tail), and had not emptied their capsid of DNA (that is, the capsid structure was not distorted and did not appear as a thin shell with a dark negatively stained centre).

There are two possible explanations for the significant number of broken phage tails. First is that the filtration of Assynt through a 0.22 µm polyvinylidene fluoride (PVDF) filter could have led to breakage (as hypothesised in **Figure 5.3**). The particle size of Assynt is relatively large for a phage, with the capsid and uncontracted tail together just over 220 nm, which is the size of the pore. If the virion is wrongly oriented and the pressure of the filtration system is high, it was hypothesised that the tails could break off, as observed in the 40% fraction TEM (**Figure 5.2C**). Second, the high pressure of the FPLC system could have led to further breakage and then concentrated the damaged particles.

PVDF matrix

PVDF 0.22 µm filter

Cross-section of PVDF matrix
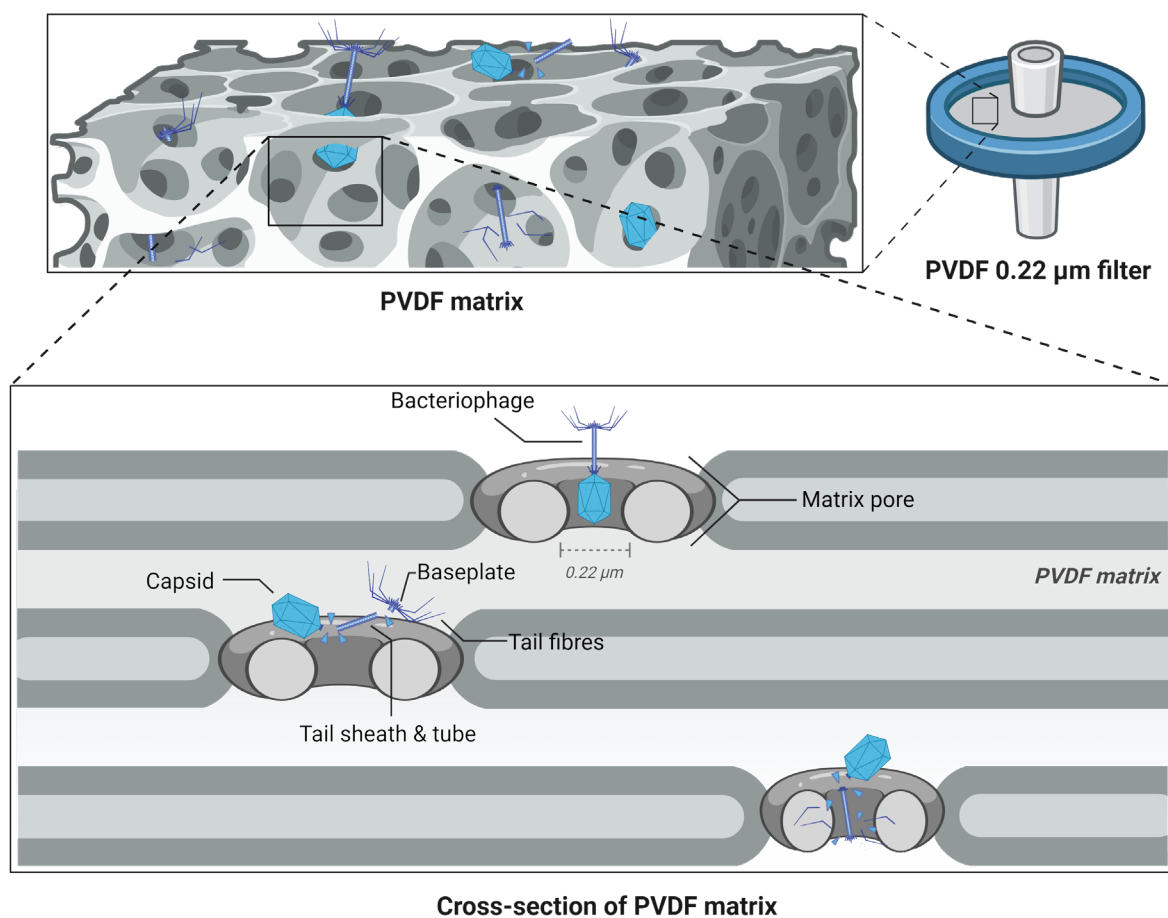
Figure 5.3. Schematic of phage Assynt filtration through 0.22 µm PVDF filter. The PVDF filter contains a matrix (top panel) that phages must pass through to be separated from larger contaminants such as bacteria and other large debris. T4-like viruses are ~0.2 µm in length and are easily broken if they do not pass through the pores head first (bottom panel). Not to scale. Figure generated in BioRender.

To determine whether phages were too strongly bound to the high-affinity (QA) column, a weak-affinity (DEAE) column was used for the second run (**Figure 5.4A-B**). A step-wise gradient was used again, with the first step at 5% elution buffer (an ionic strength of 0.1 M NaCl), which may have been able to elute more impurities. Again, two large elution peaks were observed: the 5% elution peak was collected in fractions P3 A2, P3 A3, and P3 A4; and the 40% elution peak was collected in fractions P3 B3, P3 B4, P3 B5, P3 C1, and P3 C2 (**Figure 5.4A**). This time, the two UV absorbance peaks showed longer tail ends (**Figure 5.4A**), suggesting different particle elution dynamics (Smrekar et $al.$, 2008). Additionally, there were two small peaks following 100% elution buffer, which suggested further impurities were eluted. The total phage titre that was loaded onto the column was 9.5×10^{13} in 250 mL prior to dilution (3.8×10^{11} PFU/mL). However, the total number of infective particles recovered in the 5% and 40% fractions were 2% (1.7×10^{12}), and 3.1% (2.9×10^{12}), respectively (**Figure 5.4B**, **Appendix 10.10**). Moreover, 2.0×10^{5} PFU/mL of phage was recovered from one flow through sample (P1 A1) during the second run but not the other (P1 B2), suggesting weaker binding of the phage to the weak affinity (DEAE) column, making it less suitable for phage purification (**Figure 5.4B**). As recovery of phage particles using the DEAE column was considerably lower relative to what was loaded onto the column, than recovery for the QA column, no further assays were performed.

Many areas of phage research require high-purity, high-titre phage preparations, none more so than in clinical applications (Adriaenssens et $al.$, 2012; Luong et $al.$, 2020; Sacher and Zheng, 2021). Phage amplification generally involves growing a bacterial culture and infecting the strain with the phage, and separation from bacterial cells and other contaminants by centrifugation and filtration to produce a crude phage lysate (Adriaenssens et $al.$, 2012; Bonilla et $al.$, 2016; Larsen et $al.$, 2023). Initial filtration may be followed by concentration of phage particles with PEG or TFF for example, and combined with ultrafiltration (Adriaenssens et $al.$, 2012; Larsen et $al.$, 2023). For clinical use, phage lysates must be purified to remove other bacterial debris such as endotoxin, salts, and any organic chemicals used in the amplification, concentration, and/or purification processes (Adriaenssens et $al.$, 2012; Bonilla et $al.$, 2016; Luong et $al.$, 2020; Sacher and Zheng, 2021). Moreover, these processes need to maintain phage viability.

Phage recovery during concentration and purification processes can be reduced by several orders of magnitude (Bonilla et $al.$, 2016; Larsen et $al.$, 2023; Roshankhah et $al.$, 2023). Here, rather than optimising phages using a linear elution gradient, a step-wise gradient was selected instead as generally, step-wise gradients are better suited to achieving high concentration of larger volumes of phage lysate (Smrekar et $al.$, 2008). A linear gradient may have provided a clearer indication of optimal concentration of elution buffer, with greater separation of impurities.
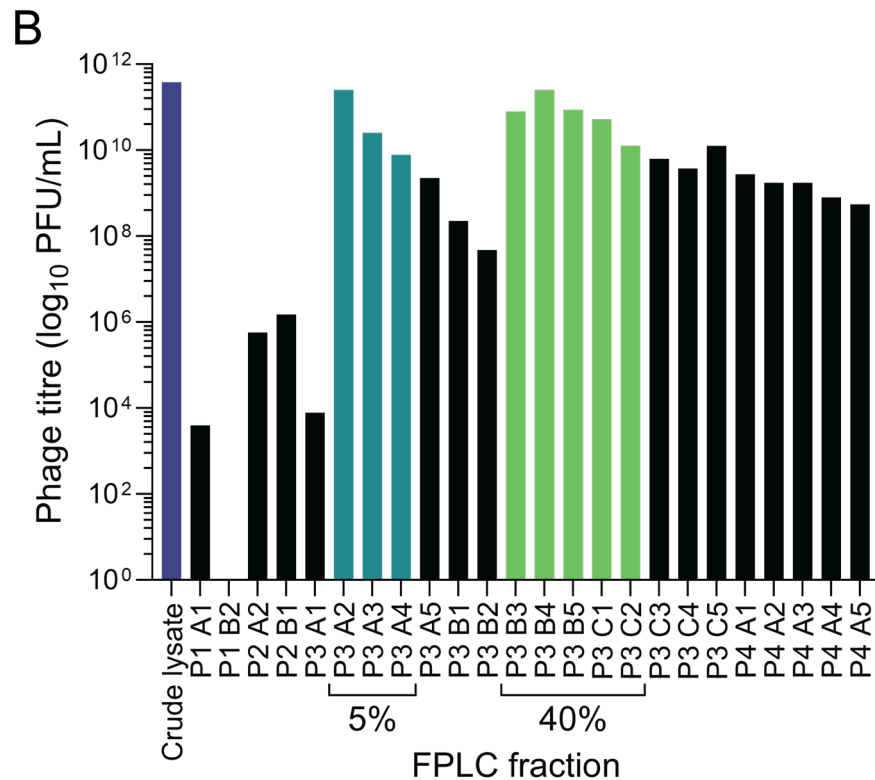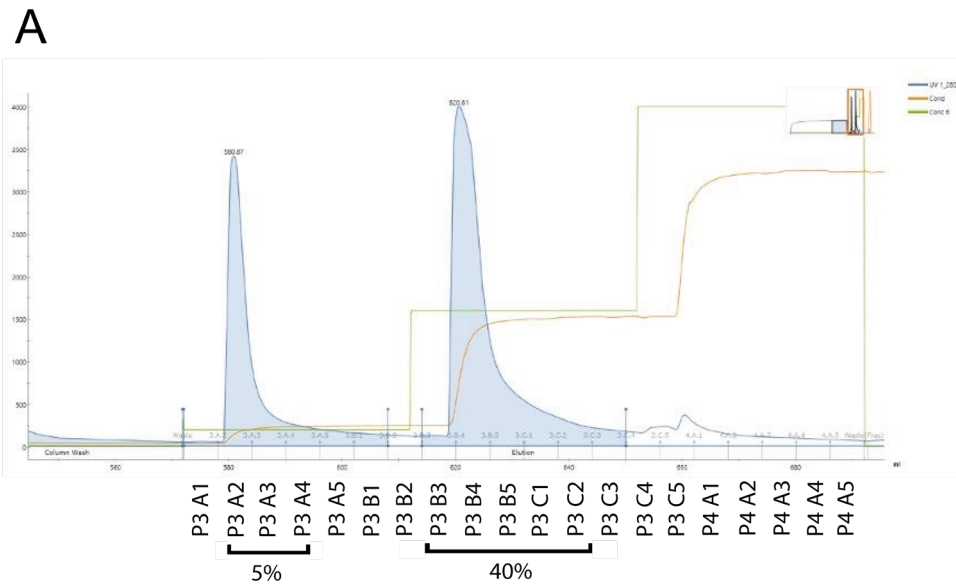
Figure 5.4. **Concentration and purification of phage Assynt by anion exchange chromatography in FPLC. (A)** Chromatogram of phage purification with a CIMmultus® DEAE monolithic low-affinity column using a step-wise gradient of elution buffer. The UV absorbance (blue line) indicated protein elution in the two peaks, which were collected and pooled as the 5% and 40% fractions, corresponding to the % concentration of NaCl in the elution buffer, respectively. The orange line represents conductivity. The green line represents the concentration of elution buffer. **(B)** Phage titres of Assynt crude lysate (purple); 5% elution peak, collected in fractions P3 A2, P3 A3, and P3 A4 (dark green); 40% elution peak, collected in fractions P3 B3, P3 B4, P3 B5, P3 C1, and P3 C2 (light green), with remaining fractions given in black. Flow through was collected in fractions P1 A1 and P1 B2.

Most phages require individual optimisation for anion exchange chromatography, which was emphasised by very poor phage recovery of 12% and 3% in their 40% fractions for the QA and DEAE columns, respectively, which was considerably less than previous reports of 30% to nearly 100% (Smrekar *et al.*, 2008; Kramberger *et al.*, 2010; Smrekar *et al.*, 2011; Adriaenssens *et al.*, 2012; Roshankhah *et al.*, 2023). For T4, Smrekar *et al.* (2008) found that phage elution occurred at 0.5 M NaCl, whereas DNA is mostly eluted at 1 M NaCl. Despite the UV absorbance profile showing only two peaks, at ~0.2 M (10%) and ~0.8 M NaCl (40%), phage was eluted in almost every fraction. Interestingly, the lack of an eluting peak during the 100% (2 M NaCl) step further suggested that little to no impurities were eluted at high ionic strength, but infective particles at relatively high concentrations (3.0×10^8 to 7.5×10^9) were still being eluted.

Phage deactivation during concentration and purification can be caused by both mechanical and chemical stress (Smrekar *et al.*, 2008; Adriaenssens *et al.*, 2012). The high concentration of broken tails in the TEM images suggested that phages were damaged either prior to loading onto the column, and also possibly by the purification system itself, leading to further reduction in recovery of infectious particles. Phage T4 demonstrates high stability in a wide range of NaCl concentrations (Smrekar *et al.*, 2008). Thus, it was presumed that Assynt, which is in the same family (*Straboviridae*) and subfamily (*Tevenvirinae*) as T4, would be relatively stable across the NaCl concentrations used during purification. However, as stability assays were not performed, chemical deactivation during purification could not be ruled out.

Whilst the lack of infective particles in the flow through suggested that the particles were not too strongly bound to the column matrix, it is possible that the NaCl concentration in the elution buffer even at 100% (2 M NaCl) was not strong enough to elute all the particles (Adriaenssens *et al.*, 2012). High ionic strength of phage lysate can result in inefficient binding of phage particles to the column (Roshankhah *et al.*, 2023), which would result in phage particles in the flow through. However, phage particles were only recovered in the flow through when using the weak-affinity DEAE column, and not the high-affinity QA column, suggesting another mechanism for low phage recovery. The volume of lysate could have impacted phage recovery: Roshankhah *et al.* (2023) found that increasing the feed volume by dilution reduced phage recovery, whereas endotoxin removal was unchanged.

The use of high-affinity anion-exchange columns combined with FPLC can concentrate and purify phages at high flow rate in a relatively short timeframe (Kramberger *et al.*, 2010; Vandenheuvel *et al.*, 2018). Moreover, the use of large bioreactors for phage amplification together with anion-exchange and FPLC allows for scalability. For example, up to eight litres of phage lysate can be produced in just four hours using large bioreactors like the models produced by Cellexus, with concentration and purification by anion-exchange and FPLC capable of processing such volumes in less than a day. To improve phage recovery and

reduce impurities, buffer exchange or ultrafiltration of the phage lysate could have reduced the ionic strength (Roshankhah *et al.*, 2023). Additionally, optimisation of purification conditions by linear gradient rather than step-wise gradient would likely improve phage recovery considerably.

5.4 Assessing endotoxin contamination of phage stocks

A limulus amebocyte lysate (LAL) assay was performed to assess endotoxin concentration in the Assynt crude phage lysate, 10% and 40% FPLC fractions at two different dilutions (**Figure 5.5**, **Appendix 10.11**).
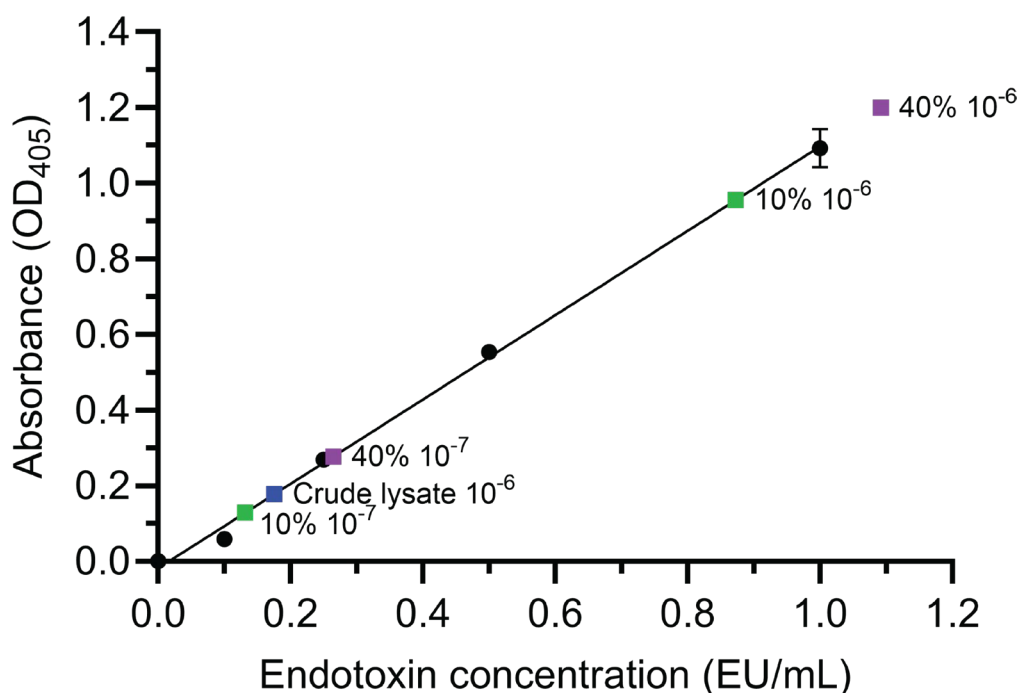
Figure 5.5. Endotoxin concentration of phage Assynt crude lysate, and pooled 10% and 40% FPLC fractions following anion-exchange with FPLC (run 1). A standard curve of 20 EU/mL endotoxin was performed with phage Assynt (diluted to 1×10^{-6} and/or 1×10^{-7}). Crude lysate, blue; 10% peak (collected in fractions P3 A2 and P3 A3) green; 40% peak (collected in fractions P3 B3, P3 B4, and P3 B5), purple. Endotoxin concentration is directly proportional to the absorbance at OD_{405}. EU/mL was extrapolated from the standard curve (black) and mean calculated for the combined dilution factors for the 10% fraction only. Coefficient of determination (r^2) = 0.9961.

Endotoxin concentrations were very high for all phage Assynt samples. The crude lysate endotoxin concentration was 1.8×10^6 EU/mL, with the 10% fraction at 1.1×10^6 EU/mL (fold change 0.63), and the 40% fraction at 2.7×10^6 EU/mL (fold change 1.50) (**Figure 5.5**, **Appendix 10.11**). Even though phage titres were lower relative to the total number of phage particles of crude lysate, the 40% fraction exhibited a 50% increase in endotoxin. This increase is unsurprising since endotoxin can co-localise with other proteins, particularly phages during concentration.

Recently, Roshankhah *et al.* (2023) found that increasing the volume of lysate by dilution increased endotoxin removal by, as yet unknown mechanisms.

Bacterial cell lysis by phages participating in the lytic cycle releases both intracellular and extracellular debris (Roshankhah *et al.*, 2023). The lipid A moiety of LPS (endotoxin), is one such major component of the outer membrane of Gram-negative bacteria, and is released upon lysis. Lipid A is highly immunogenic, with release of even small volumes capable of triggering a severe immune reaction resulting in sepsis, which can proceed to septic shock and death in patients (Gorbet and Sefton, 2005; Roshankhah *et al.*, 2023). Endotoxin is a significant contaminant in crude phage preparations whose hosts are Gram-negative bacteria. A considerable bottleneck in the application of phage preparations in clinical setting is large-scale production of high purity. Phage preparations for clinical use require a maximum of five endotoxin units per kilogram of bodyweight per hour (EU/kg/hr) for intravenous application, and < 20 EU/mL for oral application (Gorbet and Sefton, 2005; Szermer-Olearnik and Boratyński, 2015; Bonilla *et al.*, 2016).

Endotoxin contamination is ubiquitous, being found even in distilled water at concentrations of 20 EU/mL, and is thus problematic in many clinical and commercial settings (Gorbet and Sefton, 2005; Szermer-Olearnik and Boratyński, 2015). Their ability to bind cations is particularly problematic in phage production, which often use cation-based matrices to purify phages, which also have a negative charge (Gorbet and Sefton, 2005). Some endotoxin removal strategies employ organic solvents, which can result in carryover into the phage preparation, necessitating additional purification steps, often at the expense of phage titre (Szermer-Olearnik and Boratyński, 2015; Bonilla *et al.*, 2016). Moreover, many published phage concentration and purification protocols are expensive, require toxic compounds, and are time-consuming.

Like bacteriophages, endotoxin is negatively charged and can compete with them for binding to the positively charged ions within the chromatography column (Roshankhah *et al.*, 2023). Is it possible that large endotoxin aggregates could have formed complexes with intact phage virions and/or phage debris, and could have been eluted in the flow through, rather than binding to the anion exchange membrane (Roshankhah *et al.*, 2023). There were no phage particles in the flow through following testing of the FPLC fractions from the first run, but there were in the flow through of the FPLC fraction from the second run. It is possible that phage particles could have been inactivated by their binding to endotoxin aggregates and eluted in the flow through, meaning that they would not be detectable by plaque assay.

Classical methods of phage purification involve CsCl gradient density ultracentrifugation, but is not scalable. Both Luong *et al.* (2020) and Bonilla *et al.* (2016) developed standardised protocols for the production of high-quality, high-titre phage preparations. The Luong protocol combines different filtration techniques and includes an optional CsCl step,

but is still long, taking approximately 10.5 days to complete, and is not scalable. Moreover, the Bonilla protocol utilises organic compounds, which are then challenging to remove from the preparations.

The use of TFF to concentrate phages has previously been shown to effectively remove some endotoxin without compromising phage infectivity (Luong *et al.*, 2020; Sacher, 2023). Moreover, when an endotoxin removal column was used in combination with TFF, endotoxin concentration was low enough to meet the regulatory requirements of < 20 EU/mL (Luong *et al.*, 2020; Sacher, 2023). Additionally, this methodology has been shown to be suitable for standardising and upscaling phage preparations (Sacher, 2023).

5.5 Conclusions

Successful phage therapy and the development of phage-based therapeutics requires a scalable amplification method and standardised purification protocols that can be adopted into GMP. In particular, producing highly purified, high-titre phage preparations to GMP standards would aid in the regulation of phages in clinical cases in the UK (House of Commons Committee, 2024; Suleman *et al.*, 2024). To complement the biological standardisation, there should be a secure data management system in place (Lin *et al.*, 2021; Sacher *et al.*, 2022) to record the vast amount of data regarding the bacterial host, the phage, and the genomic data, including assembly methodology and quality control.

The lack of scalable and cost-effective resources has hindered high-quality high-titre phage production in the UK for clinical and compassionate use. Moreover, regulatory hurdles and the necessity of biobanking and monitoring of clinical bacterial strains and phages, as well as the secure collection and storage of data means that phage therapy in the UK is considerably behind. The data generated in this chapter show that amplification in bioreactors can yield high-titre phage lysates, however, the downstream purification is still a bottleneck as the conditions will have to be optimised for each phage individually.

CHAPTER 6: PHAGE GENOMICS

6.1 Introduction to phage genomics

Genomics is the study of the entire genome of an organism or multiple organisms in reference to their structure, function, and evolution (Owen *et al.*, 2021). Accurate genome assembly and annotation is integral to the use of phages as therapeutics (Turner *et al.*, 2021). Requirements for the therapeutic use of phages are that they do not encode any potentially harmful genes that could confer a fitness advantage to bacterial hosts; that the phage is strictly lytic; and to understand the genomic background so evolution can be tracked following propagation and/or therapeutic application (Abedon, 2017; Owen *et al.*, 2021; Turner *et al.*, 2021).

Sequencing of phage genomes is traditionally performed by short-read sequencing only due to its low costs, and low DNA input requirements (Adewale, 2020; Owen *et al.*, 2021). However, short-read sequencing and assembly can result in large structural errors, particularly within repeat regions (Wick *et al.*, 2021; Wick and Holt, 2022; Wick *et al.*, 2023). Similarly, traditional hybrid assemblies that utilise both short- and long-read sequencing as part of the assembly itself are also error-prone (Wick *et al.*, 2021; Wick and Holt, 2022; Wick *et al.*, 2023). Long-read sequencing can circumvent these issues by scaffolding of larger reads, with phage genomes capable of being captured in a single read (Owen *et al.*, 2021; Wick *et al.*, 2023). However, the per base error rate amongst long-read sequencing technologies is higher than that for short-read sequencing, and efforts to overcome this have been made with the release of the new ONT R10 chemistry (Sereika *et al.*, 2022; Cook *et al.*, 2024).

Accurate, high-quality phage genomes provide the basis of comparative genomics, leading to improved understanding of the nucleotide and proteomic composition, which contributes to the mechanistic understanding of infection biology and taxonomic classification. Together, these data can inform clinical decision-making, particularly with regard to phage cocktail design and phage-antibiotic combinations (Abedon, 2017), as well as contribute to understanding of phage diversity and ecology (Owen *et al.*, 2021; Turner *et al.*, 2021).

Here, the genomes of all phages isolated in **Chapter 4** were sequenced and assembled to produce high-quality genomes. To achieve this, a HYbrid and Poly-polish Phage Assembly (HYPPA) workflow was developed together with co-authors from across the institute[1] to produce complete, accurate, high-quality phage genomes. A full list of author contributions is available in the article cited below. HYPPA was tested and validated using ten przondoviruses as a case study. A similar hybrid poly-polish method was published for the assembly of bacterial genomes (Wick *et al.*, 2023) during the validation of HYPPA, but this is the first time such a method had been applied to phage genome assembly. Where

---

[1] Elek CKA, Brown TL, Le Viet T*, et al.* (2023). A hybrid and poly-polish workflow for the complete and accurate assembly of phage genomes: a case study of ten przondoviruses. *Microb Genom* **9**: mgen001065. All author contributions are listed in the article.

possible, HYPPA was also applied to other phages in the collection, but fully validated results are described in more detail for the przondoviruses.

Due to the volume of work contained within this chapter, a separate discussion section has been included.

6.1.1 Aims

1. Develop a workflow for generating complete, accurate, high-quality phage genomes
2. Perform assembly of all phage genomes using the new workflow
3. Perform comparative genomics for classification and taxonomic assignment of phages

6.2 Development of the HYPPA workflow for the assembly of complete phage genomes

6.2.1 Short-read-only assembly

Prior to the development of the HYPPA workflow, short-read-only and traditional hybrid assembly methods were adopted to assemble phage genomes. The most commonly used method of *de novo* phage assembly uses short-read sequence data only. Shovill was used to assemble the genomes of all phage isolates for which more than 70,000 reads of raw sequence data were generated, and SPAdes was used to assemble a single phage (Speegle) for which only 55k reads were available (**Table 6.1**). For most phages, a single contig could be recovered that represented a near-complete draft phage genome, often the largest contig in the assembly, but not for isolates Toyotomi and Fifoon, which had more fragmented assemblies (see **sections 6.2.4** and **6.2.9**, respectively). Additionally, draft genomes were recovered from short-read-only assemblies containing multiple contigs for eight phages, including Bolond (46 contigs), Speegle (56 contigs), Saitama (four contigs), Iolaire (five contigs), Kandur (three contigs), Pigeon (26 contigs), Sidders-16 (two contigs), and DeeWhy (nine contigs). Generally, the remaining contigs consisted of bacterial contamination. There appeared to be no correlation between the number of reads, read coverage and number of contigs present in each assembly.

Based on the short-read-only sequencing data, phages within the collection were preliminarily assigned a genus according to their closest database relative (see **section 6.3**). Of the 26 phages in the collection, ten were tentatively classified as belonging to the genus *Przondovirus* in the family *Autographiviridae*; two were tentatively classified as belonging to the genus *Slopekvirus*, and nine to the genus *Jiaodavirus*, both in the family *Straboviridae*; one was tentatively classified as belonging to the genus *Webervirus* in the family *Drexlerviridae*; and four were unclassified *Caudoviricetes*, of which three were siphoviruses and one was a myovirus.

Table 6.1. Short-read-only assembly details for all 26 phages in the collection.

Phage name	Assembler	No. of contigs	No. of reads	Short-read coverage [A]	Size (bp)	Size post-curation (bp)
Bolond	Shovill	46[B]	90,278	96x	176,469	176,374
Fifoon	Shovill	194	493,628	808x	*	N/A
Oda	Shovill	1	527,773	3,737x	41,540	41,641
Assynt	Shovill	1	1,244,665	2,194x	166,961	169,401
Torridon	Shovill	1	70,318	1,303x	165,906	N/A
Suilven	Shovill	2[B]	613,702	794x	169,346	N/A
Blaven	Shovill	1	1,271,918	2,223x	167,086	N/A
Toyotomi	Shovill	46	396,472	2,832x	*	41,268
Mera	Shovill	1	102,411	737x	41,315	N/A
Speegle	SPAdes	56[B]	55,213	401x	41,270	N/A
Cornelius	Shovill	1	353,905	2,608x	40,336	N/A
Tokugawa	Shovill	1	414,141	2,935x	41,312	41,413
SteelHaze	Shovill	1	921,000	5,472x	47,756	N/A
Saitama	Shovill	4[B]	1,181,498	256x	40,633	N/A
Iolaire	Shovill	5[B]	1,736,289	10,102x	48,408	48,335
Fioreun	Shovill	1	1,355,947	7,801x	48,419	N/A
Emom	Shovill	1	1,027,070	5,627x	40,692	N/A
Amrap	Shovill	1	902,433	1,846x	41,106	N/A
Kandur	Shovill	3[B]	3,526,383	5,383x	164,832	N/A
NejNej	Shovill	1	324,541	570x	165,380	N/A
Pigeon	Shovill	26[B]	2,959,068	4,338x	164,671	N/A
Sidders-16	Shovill	2[B]	106,315	178x	168,391	N/A
Eggie	Shovill	1	78,668	5,724x	164,529	N/A
Whistle	Shovill	1	1,914,368	13,511x	40,641	N/A
DeeWhy	Shovill	9[B]	2,639,052	14,532x	50,703	N/A
Dereham	Shovill	1	143,729	581x	47,376	N/A

N/A, not applicable; *, no single contig contained the entire genome: assembly was performed with subsampled reads in multiple iterations, but this did not improve the assembly quality; [A], coverage was determined for the draft assembly i.e. the largest contig; [B], genome contained within a single contig.

Several assembly methodologies were attempted to resolve the genomes of all phages and are described below. Eight phages however, could not be assembled by any other method besides short-read-only for which manual curation was incomplete, and were thus designated as draft short-read genomes: Torridon, Suilven, Blaven, Kandur, Pigeon, Sidders-16, Eggie, and Dereham (see **Table 6.2** in **section 6.2.7**).

Assessment of short read assembly quality of przondoviruses

Przondoviruses and other T7-like phages have a highly conserved genome organisation with defined chromosome ends flanked by direct terminal repeats (DTRs), making an assessment of the quality of the genome assembly simpler than for other phages in the collection. Generating complete genomes for all ten przondoviruses, including the DTRs, was initially attempted through primer walking and manual curation of the short-read-only assemblies.

From the short reads assemblies, multiple sequencing and/or assembly errors were identified in coding regions of the przondoviruses following annotation and comparative genomics (**Figure 6.1**), explained in more detail for phages Oda and Tokugawa below. Based on the comparison of the genome organisation and coding sequences (CDS) of the ten new przondoviruses with seven other members of the genus, many inconsistencies were identified that could be sequencing or assembly errors. These inconsistencies were mostly the split of CDS into multiple smaller CDSs, hypothesised to happen after the erroneous introduction of stop codons through wrong basecalls, introduction of insertions and/or deletions (indels) or incorrect assembly. Furthermore, repeat region errors, most commonly deletion of the right terminal repeat, were observed frequently (**Appendix 10.12**).

In addition to uncovering errors in these assemblies, gene clustering revealed incomplete genomes of several reference phage genomes included in the analyses (**Figure 6.1**). Specifically, the DTRs were missing in phages KP32, KPN3, and IME264. Potential errors were noted in phage KPN3 (accession MN101227), including in the DNA-directed DNA polymerase (DNAP), which was not annotated. Whilst the DNAP sequence may vary slightly between przondoviruses, the position and length of the coding sequence is highly conserved among all other przondoviruses analysed here. KPN3 was also missing the tail tubular protein A and a non-contractile tail fibre protein/spike. Whether these are genuine errors or a representation of a more divergent isolate is yet to be fully elucidated, and would require both short-read and long-read data. What is clear however, that incomplete genomes could lead to incorrect taxonomic assignments. Moreover, the presence of sequencing and/or assembly errors could infer incorrect conclusions about the nucleotide, and therefore, the proteomic structure of the phage genome. Together, these data represent important considerations when using phages in therapeutic settings, where accurate and complete genomes are necessary to understanding the evolution of the phage over time and in response to therapy.

6.2.2 Resolving defined ends and direct terminal repeats by primer walking

Due to the highly conserved nature of przondovirus genomes, T7-like phages typically have DTR regions that flank the genome and are relatively short in length (~150-400 bp). Additionally, all genes are unidirectional and show a high degree of synteny, that is the location and order of genes is conserved among the group. To facilitate a better understanding of genome organisation and synteny, the well-curated Enterobacteria phage K30 (accession HM480846) (Whitfield and Lam, 1986), was used as a reference genome for the przondoviruses in the collection. All przondovirus genomes were reoriented to start at the same start as phage K30, making the S-adenosyl-L-methionine hydrolase gene the first annotated CDS after the left DTR.
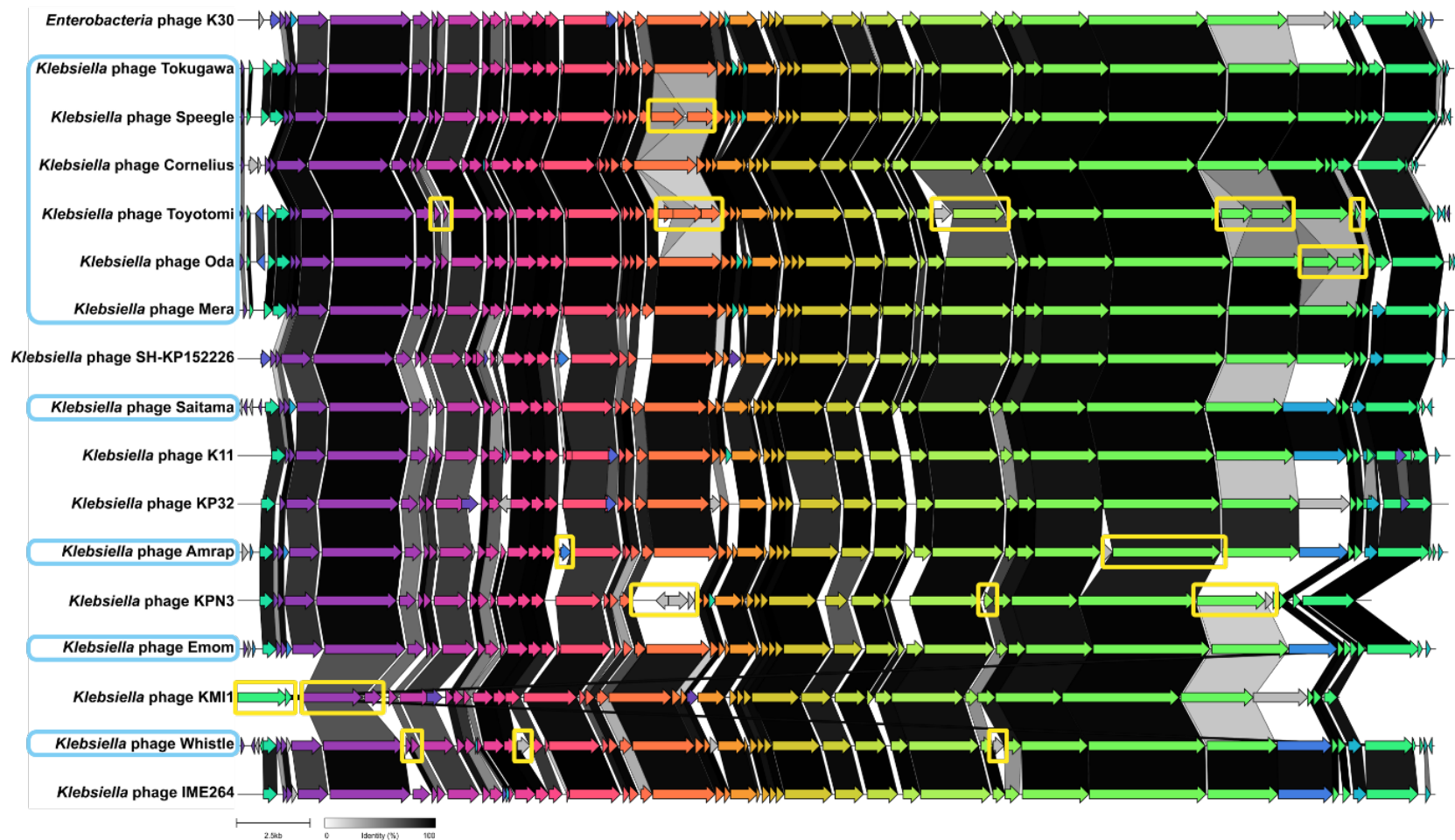
Figure 6.1. Gene cluster alignments based on protein coding sequences for przondovirus draft assemblies and a selection of related phages with potential errors. Arrows represent coding sequences and are colour coded according to similarity. Pairwise comparisons are given as links (greyscale). Coding sequences without similarity are given in grey without links. DTRs were not annotated. Przondoviruses in the collection are highlighted (blue boxes), with all genome maps representative of short-read-only data, except Toyotomi (long-read-only). Potential sequencing and/or assembly errors are given (yellow boxes). Notable errors were found in closest database relatives, including missing genes; incomplete annotations; and incorrect start sites. Figure generated with Clinker.

Next, the short-read-only assemblies were checked for DTRs, initially by manually checking the assemblies, then by using PhageTerm, which assigned incorrect phage termini to the przondoviruses and T7 because it is not suitable for use with Nextera-based sequence libraries (Garneau *et al.*, 2017). Additionally, attempts to resolve the DTRs were made by mapping the sequencing reads back to the assembly. DTRs appear as one large peak in the read mapping coverage graph as they are present in the genome twice, but only once in the assembly. The location of the peak prior to reorientation was also able to infer the location of the phage termini: if the peak was located anywhere other than at either end of the assembly, then the genome had been linearised in the middle. For example, increased reads were observed approximately 4,000 bp downstream from where the phage Oda assembly was opened, corresponding to the predicted start of the genome (**Figure 6.2**).

A problematic feature of short-read-only and traditional hybrid assemblies (those that utilise both long- and short-reads as part of the assembly) is that DTRs are incorrectly assumed by assemblers to be a sequencing artefact of a circular genome organisation, and are deleted in error. Whilst increased reads were noted for nine of the ten przondoviruses (the exception was phage Mera), all of the short-read-only assemblies had the full length of one DTR complete (which when reoriented became the left terminal repeat), but all were missing the right terminal repeat.
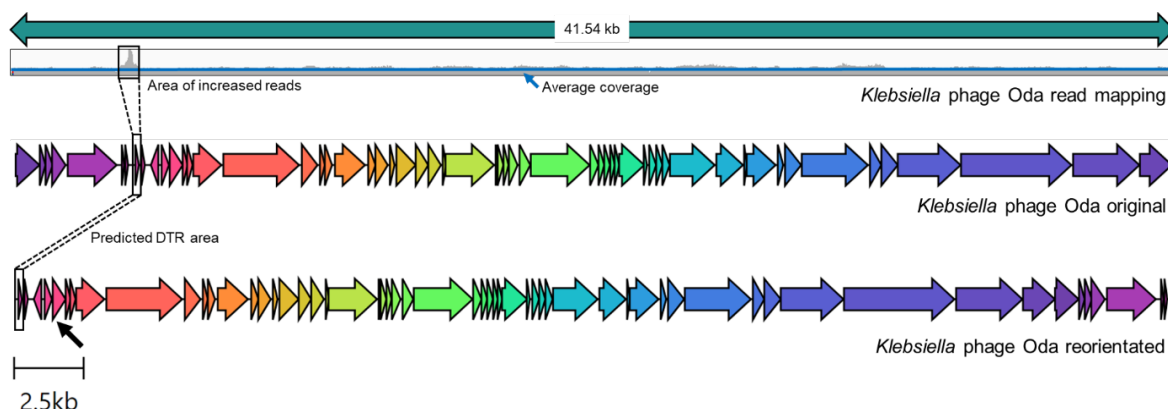
Figure 6.2. Schematic representation of the approximate location of the DTR for phage Oda. Read mapping (top panel) shows an area of increased reads located ~4,000 bp from the known genome start. Green arrow denotes short-read-only assembly length. Blue line denotes approximate average coverage. Read mapping gave an approximation of the DTR location (black boxes) and size in the original Oda assembly (middle panel) and reorientated assembly (bottom panel) based on the Enterobacteria phage K30 genome. The approximate DTR location was used to infer approximate areas for primer walking. Coding sequences for both the original and reorientated Oda assembly are represented by coloured arrows.

In an attempt to resolve the defined ends, primer walking with Sanger sequencing validation was performed to confirm the length and sequence of the DTRs for each of the ten przondoviruses. Primer walking involves Sanger sequencing by designing primers to "walk" the genome termini (**Figure 6.3A**), i.e. the sequencing reaction is performed directly on the

phage DNA with outward facing primers (Benes *et al.*, 1997). As the DTRs are present in the genome twice, typically primers are designed within and outside of the repeat region, where a clear distinction could be made from the Sanger sequencing chromatogram. Theoretically, for primers inside the repeat region, the height of the chromatogram signal drops by 50%, whilst for primers facing outwards the signal stops completely. Two or three primers were designed for each terminus, and due to the high sequence similarity, some primers were shared amongst phages.

Of the ten phages tested, primer walking was only able to resolve the DTRs for two, Tokugawa (**Figure 6.3B**) and Oda (**Figure 6.3C**). For both phages, none of the primers ended up being designed within the DTR, so a 50% decrease in the chromatograph signal was not observed, only a complete stop. Areas of the assembly where the primer walking sequences overlapped were confirmed as the DTR. For phage Tokugawa, primer walking identified that a 79 bp repeat region was incorrectly introduced during short read assembly (**Figure 6.3A).** Short repeat regions were introduced in error close to, or within the DTR regions of other przondoviruses and were initially identified through simply looking for repeats in the sequencing data, and then confirmed as either genuine or erroneous by primer walking (**Figure 6.3D**) and/or HYPPA (see **section 6.2.4**). The repeats varied in size from 13 bp to 106 bp.

Following manual curation through read mapping and primer walking with Sanger sequencing validation, a complete and accurate genome was achieved for both Oda and Tokugawa, but not for any of the other przondoviruses. Thus, attempts to resolve the remaining genomes were made using traditional hybrid and long-read-only assembly.

6.2.3 Traditional hybrid assembly for resolving phage genomes

Following unsuccessful primer walking in eight of the ten przondoviruses, whose genomes were still incomplete after short-read-only assembly, long-read sequencing was performed for the purpose of generating a traditional hybrid assembly. Traditional hybrid assembly is the use of both the short-read and the long-read data in the assembly itself, and is distinct from the hybrid assembly described in the HYPPA workflow (see **section 6.2.4**), which uses the short-read-only data for polishing only.

Initially, two different assemblers (Unicycler, which used SPAdes as the assembler, and SPAdes alone) were used to resolve the genome of Toyotomi, and then SPAdes was used to assemble the genomes of Mera and Speegle (**Appendix 10.13**). The number of contigs generated was relatively high for all the hybrid assemblies, except for the Toyotomi Unicycler assembly. However, neither assembler was able to identify both DTRs for any of the tested phages and errors found in the short-read-only assemblies persisted in the hybrid assemblies (**Appendix 10.13**).
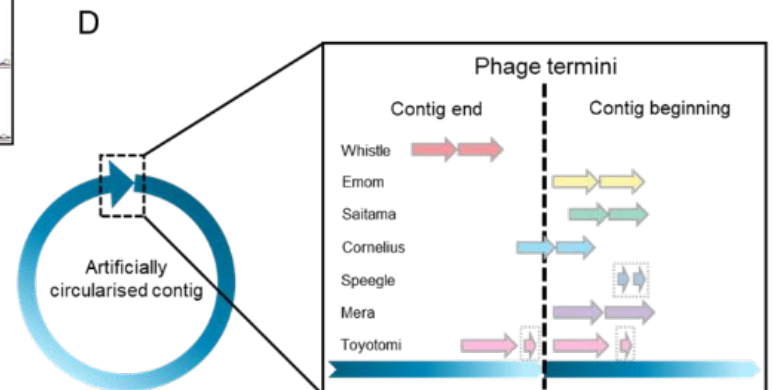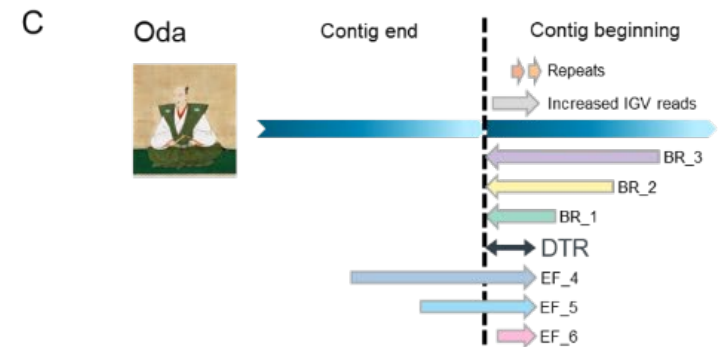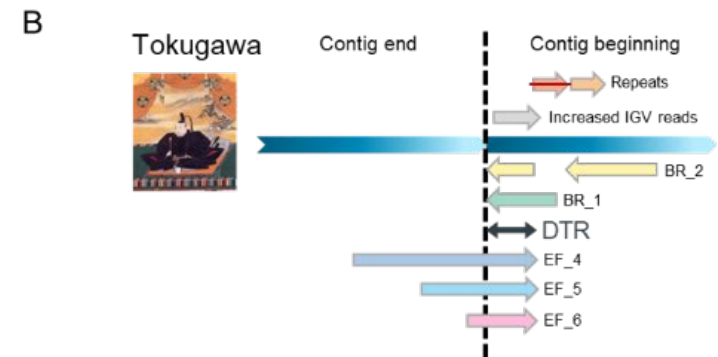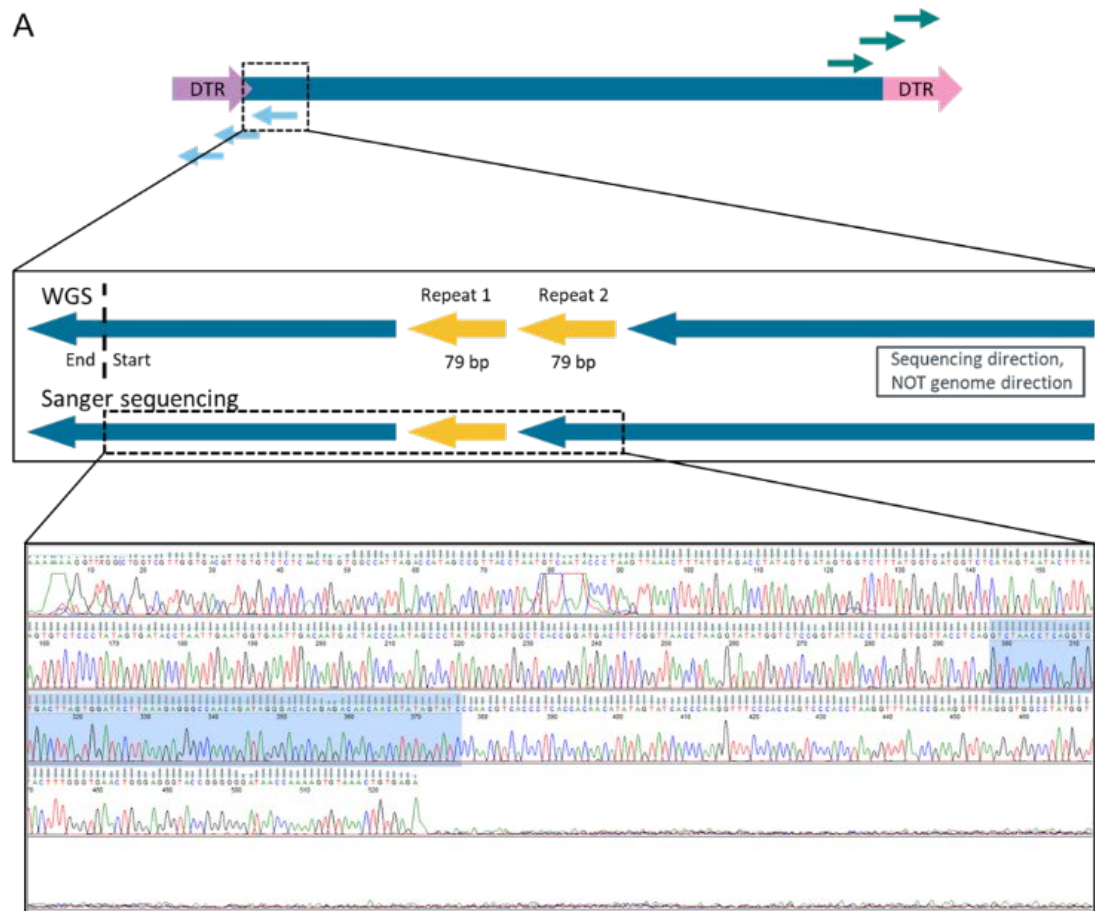
Figure 6.3. Schematic representation of primer walking and Sanger sequencing validation for phages Tokugawa and Oda. (A) Three primers were designed to "walk" the phage termini (top panel). WGS found that there was a 79 bp repeat in the short-read-only assembly of phage Tokugawa (middle panel), where only one was present in the Sanger sequencing data (middle and bottom panel) and is highlighted in the chromatograph. As the primers were all designed outside of the DTR, a 50% drop in the chromatograph was not observed. **(B)** Schematic of the primer walking and Sanger sequencing validation for phage Tokugawa. Repeats introduced in error during assembly are given in orange, with red line to demonstrate the repeat that interrupted the Sanger sequencing. The overlap of two reverse (BR) and three forward (EF) primers marked the location and size of the DTR (181 bp), and corresponded to the area of increased reads visualised in IGV (grey arrow). **(C)** Schematic of the primer walking and Sanger sequencing validation for phage Oda. Repeats (13 bp) are given in orange and determined to be genuine repeats following HYPPA. The overlap of two reverse (BR) and three forward (EF) primers marked the location and size of the DTR (181 bp), and corresponded to the area of increased reads visualised in IGV (grey arrow). **(D)** Schematic of the phage termini and the repeats found in seven other przondoviruses in the collection, varying in size from 13-106 bp (not to scale). Grey boxes denote repeats that were confirmed as genuine by HYPPA. The contig schematic has been artificially circularised (left, inset), with the termini (right) to best represent the positioning of the repeat regions.

Traditional hybrid assemblies continued to produce persistent errors, particularly in repeat regions, and if both long- and short-read libraries were constructed from different DNA extractions (Wick *et al.*, 2017; Wick *et al.*, 2017). Thus, traditional hybrid assembly was not considered robust enough to produce complete, accurate phage genomes using the ONT R9 chemistry. Only one phage genome, NejNej was assembled using traditional hybrid methods (as long-read-only assembly failed), which appeared to be slightly more accurate than the short-read-only assembly, possibly due to the newer ONT R10 chemistry. No additional manual curation was performed for phage NejNej, and it was labelled as a draft genome (see **section 6.2.7**).

6.2.4 Long-read-only assemblies

There are many different tools for assembling genomic data from long-read-only sequencing. Benchmarking on accuracy of various long-read assemblers has been performed previously (Wick and Holt, 2021), and based on these data Flye and/or Canu were used throughout this project. While short-read-only assemblers often delete DTRs, it was hypothesised that long-read sequencing would preserve the DTRs of the przondoviruses.

All przondoviruses were long-read sequenced using ONT with the R9 chemistry and assembled with Flye and/or Canu, where specified. Details of the long-read-only assemblies for the final genomes are given in **Table 6.2** in **section 6.2.7** and details of the long-read-only assemblies with alternative assemblers are given in **Appendix 10.14**.

The choice of long-read assembler used was largely dependent on the quality of the long-read sequencing data. Initially, the przondoviruses were assembled using both Flye and Canu to see if there was any difference between the two. The first assembly for all przondoviruses was performed with Flye but after assessment of the assembly quality, it was substituted for Canu for phages Oda, Cornelius, Tokugawa, Saitama, Emom, and Whistle. Flye was deemed best for genomes with high or low sequence read coverage, e.g. Toyotomi (> 117,000X), Mera (8X), Speegle (23X), and Amrap (17X) (see **Table 6.2** in **section 6.2.7**). The first assembly of both Speegle and Amrap with Flye was successful, but the second assembly with Canu, whilst also successful, resulted in massively incomplete assemblies, by > 10,000 bp (**Appendix 10.14**). The second assembly of Toyotomi with Canu failed however, possibly due to extremely high sequence read coverage. Moreover, the number of assembly errors were higher in the alternative long-read-only assemblies, even after HYPPA was applied to four (Oda, Mera, Tokugawa, and Saitama) (**Appendix 10.14**), and these long-read-only assemblies, together with the alternative long-read-only assemblies for Cornelius and Whistle were thus not used.

The most commonly encountered errors in long-read-only assemblies were in homopolymer regions, followed by indels (excluding errors within repeat regions) (**Appendix 10.12**). Interestingly, phage Toyotomi (for which a short-read-only assembly could not be generated) exhibited the most assembly errors in the long-read-only assembly, with a total of 66 errors. In contrast to the other phage assembly errors, Toyotomi exhibited more single-nucleotide errors ($n = 35$) than homopolymer errors ($n = 28$). However, homopolymer errors were particularly noticeable among coding regions for Toyotomi (**Figure 6.1**). For example, two homopolymer errors within the coding region of the DNAP lead to a double frameshift error that resulted in three protein annotations. Other homopolymer frameshift errors were found in tail fibre proteins.

Homopolymer errors are a feature of long-read-only assemblies and have been corrected effectively in bacterial genomes through short-read polishing (Wick and Holt, 2021; Wick *et al.*, 2021; Wick and Holt, 2022). However, short-read-only polishing of the long-read-only assembly of Toyotomi with Pilon not only failed to correct some persistent errors, it also introduced large deletion errors, a feature which has also been demonstrated in bacterial genomes (Wick and Holt, 2022). Moreover, the choice of assembler to use can further complicate phage assembly. Previously published literature has found that Flye is more accurate than other assemblers when used with the default settings (Chen *et al.*, 2021; Wick and Holt, 2021; Wick *et al.*, 2021). However, the data presented here demonstrated that assembly accuracy and performance may be dictated by the quality and coverage of the sequencing read data. Indeed, (Wick *et al.*, 2023) found that 200x coverage was preferable for both Illumina and ONT when assembling bacterial genomes. Moreover, the type sequencing chosen can reveal differences in the datasets, which can depend on the source of the dataset i.e. metagenomic (Cook *et al.*, 2024).

So far, short-read-only, traditional hybrid, and long-read-only assemblies had failed to resolve the genome of phage Toyotomi into a single contig. Significant manual curation was attempted to "stitch" the contigs together, followed by polishing with Pilon, which failed to correct incomplete DTRs or persistent errors, and then Racon, which failed altogether. Long-read-only assembly was able to preserve the DTRs, but errors were common and persistent, particularly in coding regions.

6.2.5 Improving przondovirus assembly: bioinformatic resolution of direct terminal repeats and assembly errors using HYPPA

Short-read-only, long-read-only, and traditional hybrid assemblies had failed to both resolve the defined ends of the przondoviruses and correct persistent assembly errors. In an attempt to generate a workflow to produce complete and accurate genomes, sequential long-read and short-read polishing steps were performed after long-read-only assembly in a workflow called HYPPA, HYbrid and Poly-polish Phage Assembly (**Figure 6.4**). HYPPA differed from traditional hybrid assembly in that it utilised a long-read-only assembly first. Thus, short-reads were only used for polishing and not for assembly.

The workflow was initially developed, tested and validated on Oda, Toyotomi and Tokugawa, three przondoviruses, which were named after the Three Unifiers of Japan, because they were considered the three unifiers of the HYPPA workflow.

Complete and accurate genomes were achieved for both Oda and Tokugawa through short-read-only assembly with manual curation, giving a solid foundation for which to test and validate the HYPPA workflow. Firstly, Oda and Tokugawa were assembled with Canu, which generated a single contig for both containing the genome at 41,761 bp and 41,532 bp, respectively. Similar to the short-read-only assembly, Oda was in the reverse orientation but unlike the short-read-only assembly the DTRs were intact. Flanking both terminal repeats however, were inverted repeats of adapter sequences and ONT barcodes, together measuring 64 bp in length that were not removed from the raw read sequencing data. Adapter and barcode sequences were identified and manually removed through read mapping.

Long-read polishing iterations were not performed as part of the assembly, and two iterations of long-read polishing was performed with Medaka. Whilst the correction itself, such as correcting single-nucleotide errors or small indels, was not always accurate, Medaka was able to correct single-nucleotide errors that were missed with short-read polishing alone. For example, a two nucleotide deletion was not corrected through short-read polishing alone, but once Medaka had corrected the error, short-read polishing was able to further identify and correct the error by removing one of the bases (**Figure 6.4**). Moreover, a single iteration of Medaka often missed single-nucleotide errors, whereas the second iteration was able to correct these errors.
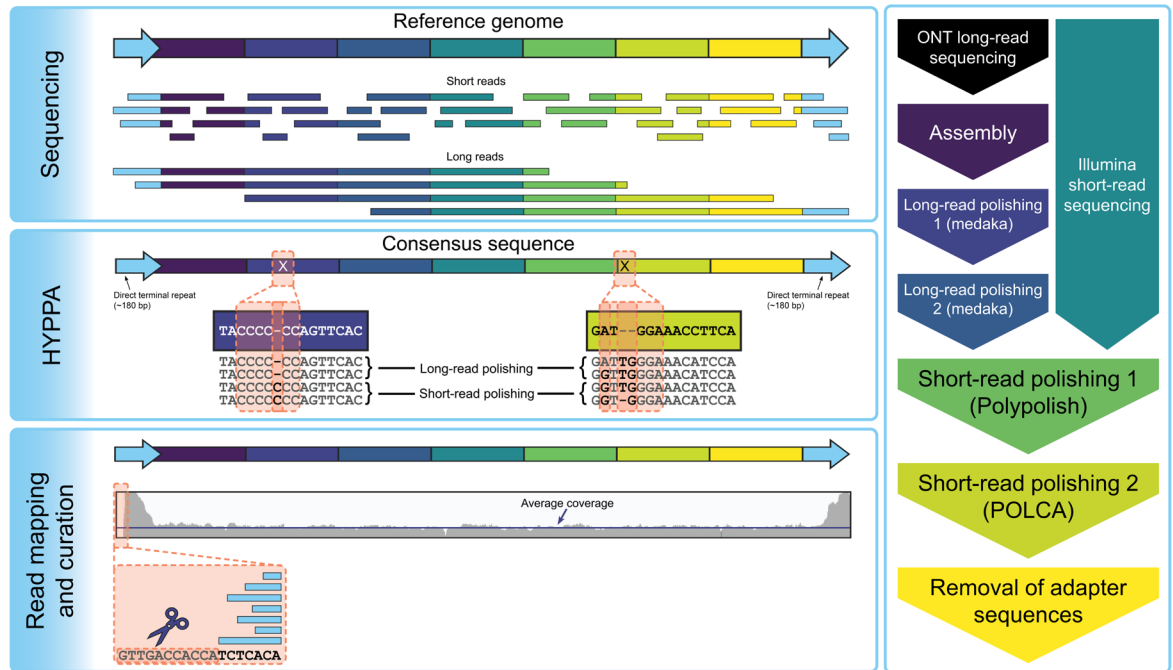
Figure 6.4. Workflow for phage genome assembly with HYPPA. Both long-read and short-read sequencing data are recommended (top panel), followed by the HYPPA workflow, which utilises long-read data to generate a consensus sequence, followed by sequential polishing steps in a specific order to achieve high-quality phage genomes (middle and right panels). Read mapping and curation to remove ONT adapters may be necessary (bottom panel).

Next, two iterations of short-read polishing were performed in a specific order. The first iteration was performed with Polypolish, which was less likely to introduce errors itself than other polishing tools. The conservative nature of Polypolish necessitated the need for long-read polishing prior, and another iteration of short-read polishing after (Wick and Holt, 2022; Wick *et al.*, 2023). Polypolish was particularly adept at correcting errors in repeat regions. The second iteration of short-read polishing was performed with POLCA, which had previously demonstrated higher accuracy in correcting errors than other polishing tools when combined with Polypolish (Wick and Holt, 2022). Both polishing tools were able to correct errors in homopolymer regions and/or indels that long-read polishing tools were unable to resolve (**Figure 6.4**). Moreover, POLCA was able to correct errors that Polypolish was unable to resolve, particularly indels. For example, POLCA was able to remove a single base insertion, that was missed by both Medaka and Polypolish (**Figure 6.4**). In some cases, polishing with POLCA was redundant as Polypolish had reduced the error count to zero.

The final curated assembly for phages Oda and Tokugawa was 41,642 bp and 41,414 bp, respectively. The DTRs were complete at 181 bp for both phages. Both HYPPA-assembled genomes showed 100% identity across the entire genome length when compared to the manually curated short-read-only assembly, thereby validating HYPPA as superior to all other assembly methods for these two phages.

Previous attempts to resolve the genome of Toyotomi into a single contig had failed. Assembly using the HYPPA workflow was able to resolve the genome of Toyotomi, resulting in a complete and accurate genome. In particular, HYPPA was able to generate a single contig, preserve the DTRs, and correct errors that persisted in all non-HYPPA assemblies.

After testing and validation, the HYPPA workflow was used for all remaining przondoviruses (see **Table 6.2** in **section 6.2.7**) resulting in complete, accurate, single-contig, DTR-containing genomes (DTRs 180-183 bp).

As an additional validation step, HYPPA was tested with both long- and short-read sequencing data constructed from the same DNA extraction for the *Escherichia* phage T7, a member of the genus *Teseptimavirus* in the family *Autographiviridae*. All sequencing data was generated for T7 as previously described for the przondoviruses. Using a well-curated T7 reference genome (accession V01146), it was found that HYPPA generated a complete and accurate genome containing zero errors, matching the reference with 100% identity across the entire genome length (data not shown). However, ONT barcode/adapter sequences at 67 bp upstream of the left terminal repeat and downstream of the right terminal repeat were found, and still had to be manually removed from the assembly. The HYPPA workflow was tested after removing the barcodes and adapters with Porechop (see **section 6.2.6**), but failed to resolve the DTRs.

In an effort to test HYPPA using sequencing data generated by other platforms, the NCBI database was searched for all przondoviruses assembled using long-read data generated by PacBio. Nine przondoviruses deposited in the database had been assembled with PacBio long-read data. However, none of the raw reads for both long-read and short-read sequencing had been made publicly-available, meaning that HYPPA could not be validated using long-read or short-read data generated by other platforms.

6.2.6 Porechop-mediated trimming of ONT adapters and/or barcodes resulted in poor-quality assemblies

ONT adapter and barcode sequences were not trimmed from the reads prior to assembling the przondovirus genomes, and were found flanking the DTRs in all przondoviruses. Initially, the adapter and barcode sequences were manually removed from the final assembly following polishing. To reduce the number of curation steps, attempts were made to remove the ONT adapter and barcode sequences with Porechop before assembly. However, Porechop-mediated trimming introduced errors, including removal of several nucleotides ranging from three to 18 bp, from the beginning of the left terminal repeat and the end of the right terminal repeat. For example, four nucleotides were removed from the beginning of the left terminal repeat, and 14 nucleotides were removed from the end of the right terminal repeat, resulting in non-homologous DTRs for phage Tokugawa when Porechop-trimmed reads were assembled with Flye (HYPPA assembly was generated with Canu). In

all cases, Porechop-trimmed reads were trialled for both the assembly and the long-read polishing. Assembly with the Porechop-trimmed reads and polishing with the raw ONT reads resulted in no difference between the two assemblies, with errors still persisting for phages Tokugawa and Cornelius.

In some cases, Porechop-mediated trimming prior to assembly resulted in failure of the genome to assemble into a single contig, or even at all, depending on the assembler used. For example, when Porechop-mediated trimmed reads were used, Tokugawa failed to assemble at all with Canu as the assembler. Thus, Porechop-mediated trimming resulted in poorer-quality genome assembly, and manual removal of ONT adapter and/or barcode sequences was necessary. When using Porechop-trimmed reads, both Mera and Speegle were assembled into multiple contigs that required significant manual curation to generate an assembly in a single contigs, where significant errors introduced during assembly could not be corrected through polishing.

Preliminary investigation of other HYPPA-assembled phages, including those in the family *Straboviridae*, unclassified siphoviruses, and unclassified myoviruses found no ONT adapter or barcode sequences. A lack of ONT adapter or barcode sequences is usually because the genomes are circularly permuted, and the assembler is able to correctly trim the adapters from the assemblies.

6.2.7 HYPPA-mediated assembly for other phage genomes

The HYPPA workflow has successfully generated complete and accurate genomes for przondoviruses, where the validation was simplified due to the highly conserved nature of the genome organisation in this particular genus and the availability of high-quality reference genomes. HYPPA was then applied to all other phages in the collection, many of which belonged to different genera, and had different genome composition and organisation.

A total of 16 phages were long-read sequenced, all with the R9 chemistry. During the R9 run, there were issues with incorrect barcoding, and cross-contamination due to impartial activation of the adapter ligase, which meant long-read sequencing for the 16 phages was delayed. Subsequently, long-read sequencing was repeated for ten of the 16 phages with the new R10 chemistry but eight failed to assemble. Assembly failure was likely due to poor quality sequencing data: all assemblies using Flye were across multiple contigs; the read length N50 was low for all except Blaven; and the number of reads was low for many (**Appendix 10.15**). Additionally, possible fragmentation of DNA during DNA extraction could have contributed to poor ONT sequencing in some cases.

Only two phages were assembled using the R10 chemistry, Assynt and NejNej (discussed below), using the HYPPA workflow and traditional hybrid with Unicycler, respectively.

However, final genome verification was not performed due to time constraints, and both were designated draft genomes.

A single phage in the genus *Jiaodavirus*, NejNej was assembled by traditional hybrid methods using the new R10 chemistry, because long-read-only assembly and subsequent HYPPA workflow generated fractured assemblies across multiple contigs with both Canu (data not shown) and Flye (**Appendix 10.14**). Mean read quality scores for the long-read data were within the generally accepted threshold of 8-12 (NejNej mean read quality = 10.6). However, the read length N50 was relatively short (637 bp), and the coverage was relatively low (50x), but still comparable to other phages that assembled successfully with HYPPA (**Table 6.2**, **Appendix 10.14**). No further curation was performed for this phage due to time constraints.

Eight phages could not be assembled using the HYPPA workflow, likely due to poor quality long-read sequencing data, and consisted of short-read-only assemblies (Torridon, Suilven, Blaven, Kandur, Pigeon, Sidders-16, Eggie, and Dereham) (**Table 6.2**, **Appendix 10.15**). Thus, all were designated draft genomes.

Six remaining phages were assembled using the HYPPA workflow with the R9 chemistry (Bolond, Fifoon, SteelHaze, Iolaire, Fioreun, and DeeWhy), of which two had fractured assemblies that were difficult to resolve and are discussed separately below. Due to time constraints, the completeness and accuracy of Fifoon, SteelHaze, and DeeWhy assemblies could not be fully determined, and were also designated as draft genomes.

The complete and draft genomes of all 26 phages are given in **Table 6.2** along with sequencing metadata.

The HYPPA assemblies described below generally follow a similar format: the name of the phage according to their lab code and the version according to the number of times they have been sequenced (i.e. CE1v3), followed by the assembler used (Flye), with a version number given to a particular step in the HYPPA workflow:

- v1, assembly without polishing
- v2, assembly plus two iterations of long-read polishing
- v3, assembly plus two iterations of long-read polishing and two iterations of short-read polishing i.e. completed HYPPA workflow

Table 6.2. Assembly details for all 26 phages in the collection.

Phage lab code	Phage name	Assembly type	Assembler	No. contigs	No. of ONT/SR reads	ONT Read length N50	Longest ONT read (bp)	ONT/SR read coverage	Size pre-polishing (bp)	Size post-polishing* (bp)	Final assembly (bp)
CE1	Bolond	HYPPA	Flye	3[a]	20,000	5,615	52,679	1,107x	176,357	176,374	176,374
CE2	Fifoon	HYPPA	Flye	1	360,654	4,510	75,469	2,457x	175,966	176,192	-[c]
CE3	Oda	HYPPA	Canu	1	1,825	12,634	41,869	146x	41,761	41,769	41,642
CE4	Assynt	HYPPA	Flye	1	9,130	579	54,472	30x	168,971	169,401	-[c]
CE5	Torridon	SR	Shovill	1	70,318	-	-	1,303x	165,906	-	-[c]
CE6	Suilven	SR	Shovill	2[a]	613,702	-	-	794x	169,346	-	-[c]
CE7	Blaven	SR	Shovill	1	127,032	-	-	2,223x	167,086	-	-[c]
CE8	Toyotomi	HYPPA	Flye	1	365,482	39,443	106,111	117,981x	41,249	41,286	41,268
CE9	Mera	HYPPA	Flye	3[a]	13,252	12,497	58,219	8x	41,454	40,619	41,400
CE10	Speegle	HYPPA	Flye	1	47,260	12,610	81,712	23x	41,394	41,395	41,395
CE11	Cornelius	HYPPA	Canu	1	5,719	1,388	47,884	118x	40,555	40,564	40,437
CE12	Tokugawa	HYPPA	Canu	1	1,018	19,093	41,581	92x	41,532	41,542	41,414
CE14	SteelHaze	HYPPA	Flye	5[a]	1,147,510	885	78,891	11,853x	47,655	47,672	47,677[c]
CE15	Saitama	HYPPA	Canu	1	19,217	2,597	43,752	661x	46,600	46,613	40,741
CE18	Iolaire	HYPPA	Canu	1	152,000	9,312	99,092	6,649x	96,642[b]	50,894	48,335
CE19	Fioreun	HYPPA	Flye	10[a]	136,000	5,192	94,930	3,013x	96,672[b]	48,347	48,347
CE20	Emom	HYPPA	Canu	1	2,683	4,862	41,521	52x	40,903	40,915	40,788
CE22	Amrap	HYPPA	Flye	2[a]	30,181	717	40,963	27x	41,328	41,331	41,209
CE24	Kandur	SR	Shovill	3[a]	3,526,383	-	-	5,383x	164,832	-	-[c]
CE25	NejNej	Hybrid	Unicycler	1	23,516	637	37,626	50x	165,293	165,293	-[c]
CE26	Pigeon	SR	Shovill	26[a]	2,959,068	-	-	4,338x	164,671	-	-[c]
CE27	Sidders-16	SR	Shovill	2[a]	106,315	-	-	178x	168,391	-	-[c]
CE28	Eggie	SR	Shovill	1	78,668	-	-	5,724x	164,529	-	-[c]
CE30	Whistle	HYPPA	Canu	1	9,239	1,614	46,197	200x	40,855	40,862	40,735
CE32	DeeWhy	HYPPA	Flye	1	8,000	16,720	85,042	320x	101,248[b]	50,630	-[c]

143

CE34	Dereham	SR	Shovill	1	143,729	-	-	581x	47,376	-	-[c]

*, post-polishing was 2 x long-read and 2 x short-read polishing, as per the HYPPA workflow; [a], genome contained within a single contig; [b], the original assembly was double the size, so the genome was separated into two contigs prior to polishing; SR, short-read-only; -, no data. Number of reads for HYPPA and hybrid assemblies are given as number of long-reads only, for short-read metadata, see **Table 6.1**. Final assembly included any post-polishing curation that was necessary; [c], currently designated draft genomes.

HYPPA-mediated assembly for phage Bolond

Phage Bolond (lab code CE1) was assembled without issue with HYPPA using Flye as the assembler (CE1v3_flye_v1, 176,357 bp) with two iterations of long-read polishing (CE1v3_flye_v2, 176,371 bp) followed by two iterations of short-read polishing (CE1v3_flye_v3, 176,374 bp) (**Table 6.2**). All Bolond assemblies were in an orientation that was considered reverse compared to reference genomes and linearised at random locations in the genome. Thus, the reverse complement and all assemblies were reorientated to the same starting position, which was arbitrarily selected as the starting sequence of a tail fibre protein proximal subunit. For this phage, and all other phages, annotations were performed with Pharokka using PHANOTATE for gene calling and the PHROGs database for functional annotation.

When assembled with Canu, the assembly was approximately 110% longer than expected (195,877 bp). This may be an artefact of long-read sequencing and assembly of genomes that are circularly permuted. Moreover, the Canu assembly contained more assembly errors in tail fibre proteins, topoisomerase, DNAP, and baseplate proteins, suggesting that Flye was the more accurate assembler for this particular phage (**Figure 6.5**). When compared to the HYPPA assembly (CE1v3_flye_v3), the only major difference in the short-read-only assembly (CE1v1_shovill, 176,374 bp) was a 95 bp insertion in the latter. Through read mapping, the 95 bp insertion was identified as having been introduced in error, and was easily observed in the gene clustering (**Figure 6.5**). Thus, the HYPPA assembly for Bolond was determined as the most complete and accurate, and was taken forward for further analyses.

HYPPA-mediated assembly for phages Iolaire and Fioreun

Like all the other phages, Iolaire and Fioreun (lab code CE18 and CE19, respectively) were initially assembled using short-read-only data, and Iolaire underwent significant manual curation of the short-read-only assembly to correct assembly errors prior to development of the HYPPA workflow. Thus, both the original and curated short-read-only assemblies were used for comparisons.

Following long-read-only assembly with Flye, both Iolaire and Fioreun assembled into a single contig each, but the genome was present twice, resulting in a contig that was double the size (**Table 6.2**). The double genome within the single contig was extracted as genome1 and genome2 for Iolaire, and the HYPPA workflow was applied to both genome1 and genome2. A multisequence alignment (MSA) was performed between the genome1 and genome2 following HYPPA to compare any potential differences between the two. Genome1 and genome2 were identical, however. The process of separating and determining differences between genome1 and genome2 was repeated for phage Fioreun, and again, both genomes were identical. Genome1 and genome2 for both Iolaire and Fioreun following HYPPA were included in the gene clustering analysis (**Figure 6.6**).

Figure 6.5. Gene cluster alignments based on protein coding sequences for phage Bolond (lab code CE1) long-read-only, short-read-only, and HYPPA assemblies. Arrows represent coding sequences and are colour coded according to similarity. Pairwise comparisons are given as links (greyscale). Coding sequences without similarity are given in grey without links. Default display order is based on cluster similarity. HYPPA assembly used for further analyses was the CE1v3_flye_v3. Figure generated with Clinker.

Similarly to other assembly workflows described for Bolond, a long-read-only assembly was performed using Flye and/or Canu, followed by the HYPPA workflow, and the assemblies at various stages of HYPPA were used in the comparisons, along with the short-read-only assembly. Despite showing high sequence similarity across the genome length, Iolaire and Fioreun were analysed separately for assembly analyses.

For Iolaire, major differences were observed in the original short-read-only assembly (CE18v1_shovill) and the unpolished long-read-only assembly (CE18v2_flye_v1_genome1), mostly between the virion assembly proteins, notably tail-associated proteins and the terminase, exonuclease, holin, Rz-like proteins, DNA polymerase/primase, as well as some hypothetical proteins (**Figure 6.6A**).

Sixteen mismatches were identified in the unpolished long-read-only assembly, all of which were errors in homopolymer regions. The short-read-only assembly contained only one major error when compared to both final HYPPA assemblies (CE18v2_flye_v3_genome1 and genome2) and the curated short-read assembly, which was a 73 bp repeat region in the tail protein, that was introduced in error during assembly.

No mismatches were observed between both HYPPA assemblies (genome1 and genome2), the curated short-read-only assembly (CE18v1_shovill_curated), and the long-read polished assembly (CE18v2_flye_v2_genome1), suggesting that long-read assembly was able to correct structural errors introduced by the short-read data. Interestingly, the long-read polishing with Medaka was able to correct all homopolymer errors, which are usually only corrected following short-read polishing. Thus, no differences between the HYPPA assemblies and the long-read polished assembly were observed.

Similar types of errors were observed for Fioreun, with clear differences between the short-read-only assembly (CE19v1_shovill), the unpolished long-read-only assembly (CE19v2_flye_v1_genome1), and the polished assemblies (**Figure 6.6B**). Notably, the short-read-only assembly contained errors in tail-associated proteins, holin/antiholin, Rz-like spanin, and DNA polymerase/primase.

There were 13 mismatches identified in the unpolished long-read-only assembly of Fioreun when compared to the final HYPPA assemblies (CE19v2_flye_v3_genome1 and CE19v2_flye_v3_genome2) and the short-read-only assembly. The short-read-only assembly also contained a 73 bp repeat in the tail length tape measure protein, and one single-nucleotide error in the first protein that were introduced in error during assembly and corrected following HYPPA. Thus, final complete and accurate genomes for used for further analyses for Iolaire and Fioreun were the HYPPA assemblies, CE18v2_flye_v3_genome1 and CE19v2_flye_v3_genome1, respectively.
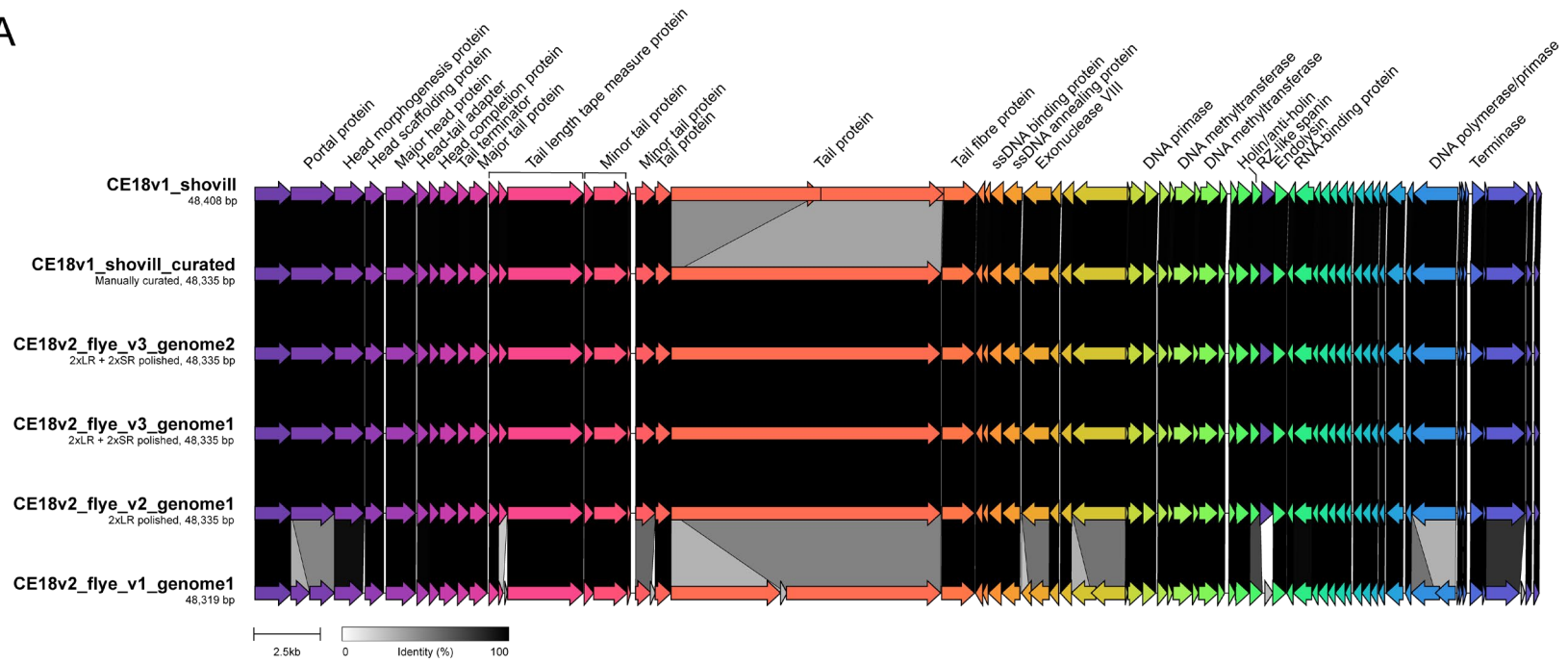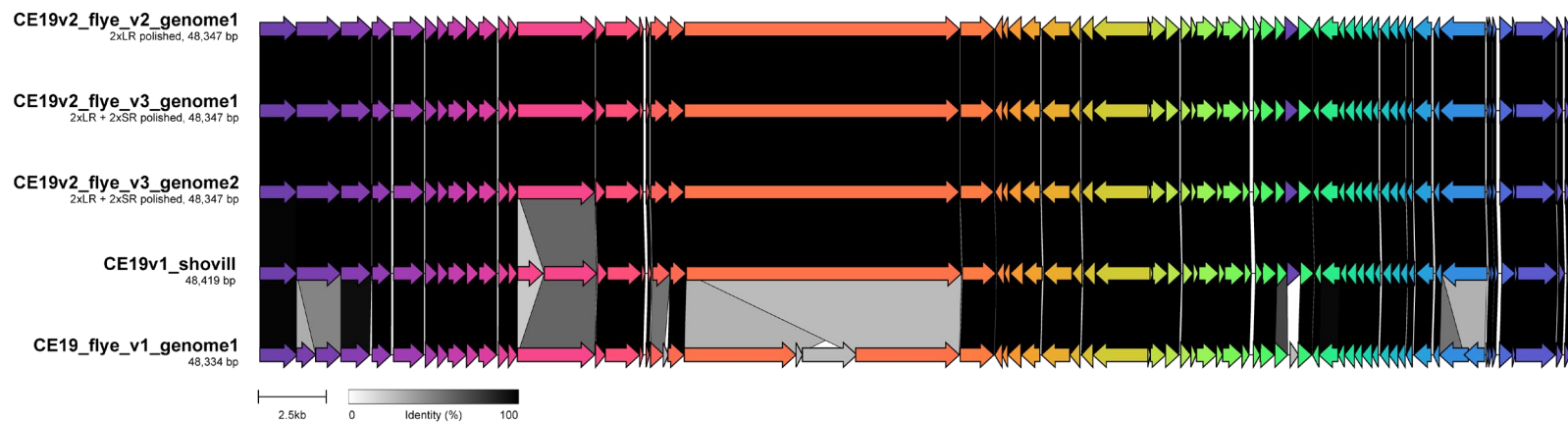
Figure 6.6. Gene cluster alignments based on protein coding sequences for assemblies of phages Iolaire and Fioreun. (A) Long-read-only, short-read-only, and HYPPA assemblies at various stages for the Iolaire (lab code CE18) assembly extracted as genome1. Only the final HYPPA assembly is shown for genome2. Both the original short-read and curated short-read assemblies are given. **(B)** Long-read-only, short-read-only, and HYPPA assemblies at various stages for the Fioreun (lab code CE19) assembly extracted as genome1. Only the final HYPPA assembly is shown for genome 2. Arrows represent coding sequences and are colour coded according to similarity. Pairwise comparisons are given as links (greyscale). Coding sequences without similarity are given in grey without links. HYPPA assemblies used for further analyses for Iolaire and Fioreun were the CE18v2_flye_v3_genome1 and CE19v2_flye_v3_genome1, respectively. Figure generated with Clinker.

HYPPA-mediated assembly for phage DeeWhy

Similarly to Iolaire and Fioreun, the DeeWhy genome was present twice in a single contig of the assembly. An MSA between the two genomes (genome1 and genome2) found no differences, and only the second copy, CE32v2_flye_v1_genome2 was used for further analyses.

Six mismatches were observed between the long-read-only assembly (CE32v2_flye_v1_genome2) and the HYPPA assembly (CE32v2_flye_v3_genome2) (**Figure 6.7**), as for Iolaire and Fioreun, all of which were in homopolymer regions. Mostly, the mismatches did not result in significant errors in the genome, besides a frameshift error in one hypothetical protein. Again, a 73 bp insertion within the DNA helicase protein was found in the short-read-only assembly and was introduced in error, resulting in a clear frameshift error and two annotations within a hypothetical protein.

The final genome used in further analyses was CE32v2_flye_v3_genome2, but as no further checking of the assembly was performed to correct any potential errors, this genome was still designated a draft genome at the time of writing.

6.2.8 HYPPA-mediated assembly of phage Assynt with the new R10 chemistry

The final phage to be successfully assembled using the HYPPA workflow was phage Assynt, using the new R10 chemistry to generate long-read data. Assembly using the R9 chemistry failed for this particular phage, due to incorrect barcoding, and cross-contamination between phages due to impartial activation of the adapter ligase. Thus, a direct comparison between the assembly quality using the R9 and R10 chemistries was not possible.

Additionally, a traditional hybrid assembly with the R10 chemistry was used in the assembly comparisons analyses, which had also undergone the same polishing iterations as described in HYPPA.
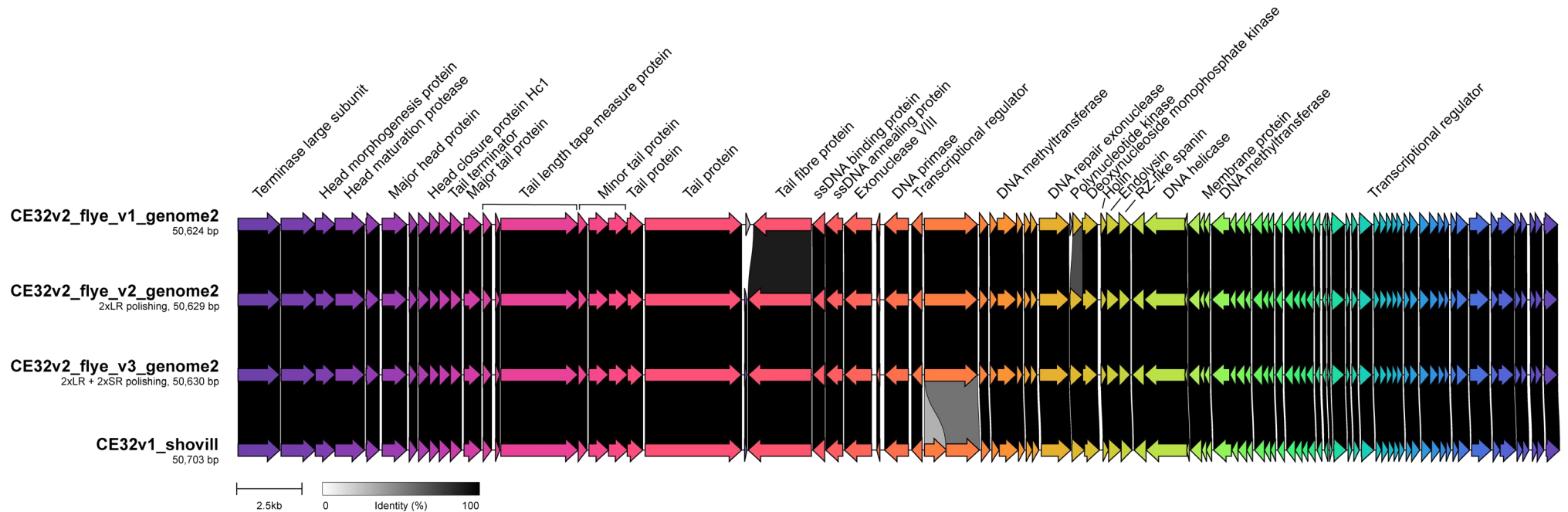
Figure 6.7. Gene cluster alignments based on protein coding sequences for phage DeeWhy (lab code CE32) long-read-only, short-read-only, and HYPPA assemblies. Arrows represent coding sequences and are colour coded according to similarity. Pairwise comparisons are given as links (greyscale). Coding sequences without similarity are given in grey without links. Default display order is based on cluster similarity. Figure generated with Clinker.

Gene clustering was performed between the original short-read-only assembly (CE4v1_shovill), the long-read-only (CE4v2_flye_r10_v1) and long-read polished (CE4v2_flye_r10_v2) assemblies that were the prerequisites to HYPPA, the HYPPA assembly (CE4v2_flye_r10_v3) and the traditional hybrid assembly with HYPPA polishing (CE4v3_unicycler_r10_v3) (**Figure 6.8**). Large structural differences were observed in the long-read-only and long-read polished assemblies compared to the others.

A higher number of errors were present in the long-read-only and long-read polished assemblies when compared to the short-read-only, traditional hybrid, and HYPPA assemblies. It is difficult to determine whether the R10 chemistry contributed to these differences as there was no way of comparing them to an R9 run. Minor differences were observed in coding regions between the HYPPA assembly and the traditional hybrid assembly, including the presence of genes not found in the latter, as well as differences in genome length of 670 bp. Whilst the short-read-only assembly was mostly similar to the HYPPA and hybrid assemblies in its coding sequences, its genome length was 2,263 bp and 1,593 bp shorter, respectively. These differences were most often due to large structural errors and/or large deletions in the short-read-only assembly. However, the legitimacy of these errors could not be verified, and the final draft genome used in further analyses was the CE4v2_flye_r10_v3 (HYPPA) assembly.

Together, these data show that careful curation and checking of genomes is integral prior to their use in further downstream analyses. If using the short-read-only assembly in comparative genomics for example, it could lead to incorrect conclusions being made regarding protein structure and function, and taxonomic classification.

6.2.9 Resolving genomes of fractured or problematic HYPPA assemblies

Here, two phages were assembled with the HYPPA workflow with varying degrees of success: Fifoon in the genus *Slopekvirus*; and SteelHaze, an unclassified myovirus.

Resolving the genome of phage Fifoon

Similarly to phage Toyotomi, Fifoon (lab code CE2) could not be assembled into a single contig with the short-read-only data. Fifoon's closest relative in the collection was phage Bolond (98.76% identity over 98% genome length according to BLASTn), which assembled without issue. Moreover, three runs of short-read sequencing and assembly for Fifoon were performed as it was initially assumed to be a sequencing issue rather than an assembly issue. This was followed by one run of long-read sequencing and multiple assembly attempts (**Table 6.3**, **Figure 6.9**).
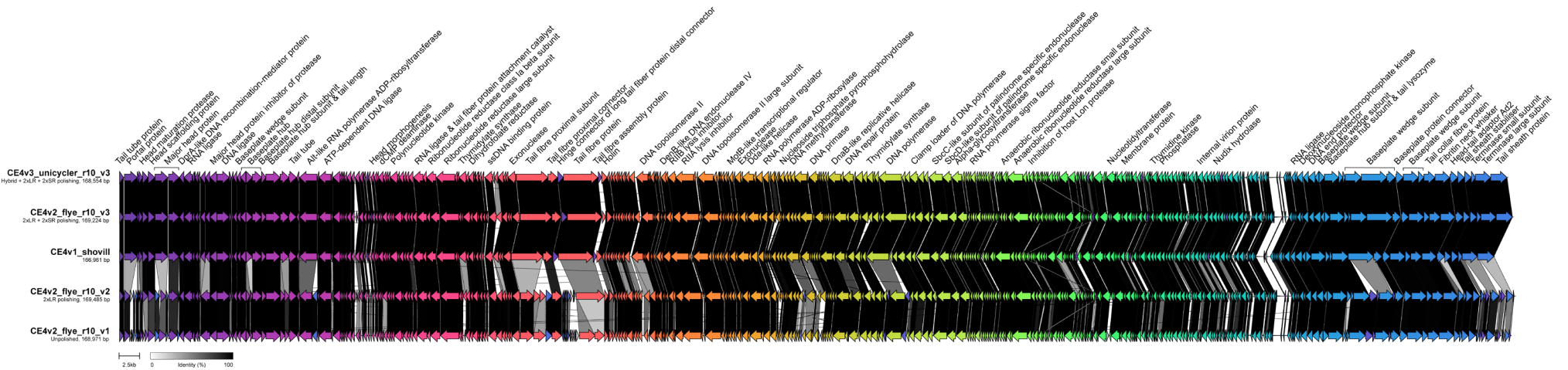
Figure 6.8. Gene cluster alignments based on protein coding sequences for phage Assynt (lab code CE4) long-read-only, short-read-only, traditional hybrid, and HYPPA assemblies. Arrows represent coding sequences and are colour coded according to similarity. Pairwise comparisons are given as links (greyscale). Coding sequences without similarity are given in grey without links. Default display order is based on cluster similarity. HYPPA assembly used for further analyses was the CE4v2_flye_r10_v3. Figure generated with Clinker.

Eighteen assemblies were generated and analysed, but do not represent the total number of assemblies that were generated through trial and error. Many of the assemblies are variations of one another, such as those at different stages of HYPPA, different iterations of long-read-only assemblies, or different combinations of polishing, and thus ten of the 18 assemblies were used for further analyses, with the assembler given in the assembly name (**Table 6.3**).

Table 6.3. Assembly comparisons for phage Fifoon (lab code CE2) following long-read-only assembly with Flye or Canu and various stages of the HYPPA workflow.

Phage name	Assembly stage	Size (bp)
CE2_flye_v1	LR-only	175,998
CE2_flye_v7	HYPPA + HYPPA	176,264
CE2v1_flye_stitched	LR-only + manually stitched	176,377
CE2v1_flye_stitched_polished	LR-only + manually stitched + SR polishing[a]	176,536
CE2v3_flye_v1_contig1	LR-only	175,966
CE2v3_flye_v1_contig2	LR-only	176,312
CE2v4_canu_v2	LR-only + LR polishing	221,603
CE2v5_flye_v3.1	HYPPA	176,229
CE2v5_flye_v3.2	HYPPA + SR polishing[a] + variant calling	176,192
CE2v5_flye_v4	HYPPA (4x LR polishing)[b]	176,346

LR, long-read; SR, short-read. [a], SR polishing performed separately with Pilon; [b], 4x LR polishing performed with Medaka as part of the HYPPA workflow.

Through trial and error, the fewest contigs Fifoon could be assembled into using short-read-only data were 194, with many other iterations being more than 300 contigs. Long-read-only assembly with Flye produced better results when the genome was assembled with Canu. For example, Canu assemblies were each approximately 45,000 bp longer than their Flye counterparts (~125% longer). An erroneous area in the middle of the genome appeared to be responsible for the difference in genome length, and did not appear to be a result of linearisation of the genome by the assembler. The erroneous area showed 99% identity across 99% of the genome with other slopekviruses. After some preliminary analyses, the assembly contained many genomic repeats, but were not in a consistent order. For example, one repeat was 282 bp, another was 41 bp, and another was 33 bp, with regions between some repeats that appeared to be part of the genuine assembly. Gene clustering confirmed that a large proportion of the genome was repeated (**Figure 6.10**). A possible explanation could be that phages in the family *Straboviridae* have genomes that are circularly permuted, where the genome can be assembled as a concatemer, and the final assembly can include more than 100% of the genome length and in some cases even multiple copies in a row.

Similarly to phage Bolond, all Fifoon assemblies had to be reorientated to the same starting position, which the starting sequence of a tail fibre protein proximal subunit.
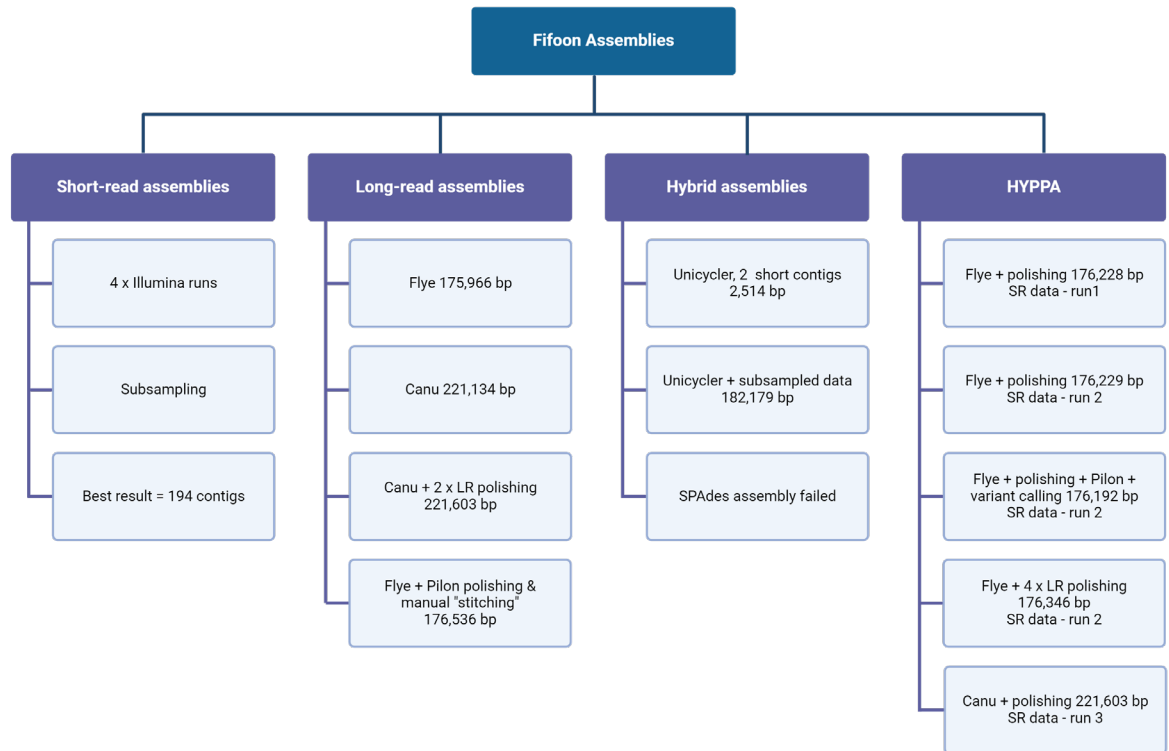
Figure 6.9. Flow diagram for a selection of phage Fifoon assemblies. Three iterations of short-read sequencing were performed. Long-read-only assembly was most successful using Flye, but Pilon-mediated polishing required significant manual curation. Hybrid assemblies used two different SR sequencing data. HYPPA assemblies used three different SR sequencing data. Not all assemblies listed here were used in further analyses. SR, short-read; LR, long-read.

Three different long-read-only assemblies without further polishing or curation were generated (CE2_flye_v1, CE2v3_flye_v1_contig1, CE2v3_flye_v1_contig2), which resulted in three different genome lengths (**Table 6.3**) with differences between certain genes according to the gene clustering, such as in tail fibre proteins, DNAP, baseplate hub, and baseplate wedge proteins (**Figure 6.10**). Interestingly, all three assemblies were generated from the same set of long-read data, using the same version of the same assembler, with contig1 and contig2 assembled around the same date.

Gene clustering revealed incomplete assemblies for all those analysed, of which there was no clear candidate to take forward for comparative genomics as all contained potential errors. Indeed, no assembly appeared to be without errors, despite attempting different polishing programs in different combinations, with different short-read data (**Table 6.3**, **Figure 6.10**). Obvious errors were observed in DNA replication proteins (DNAP, primases, topoisomerases *etc.*) and virion assembly proteins (tail fibre proteins, baseplate proteins, portal proteins *etc.*). However, gene clustering revealed fewer potential errors in the assembly of ce2v5_flye_v3.2, particularly in the tail fibre proteins and this draft genome was used for further analyses.
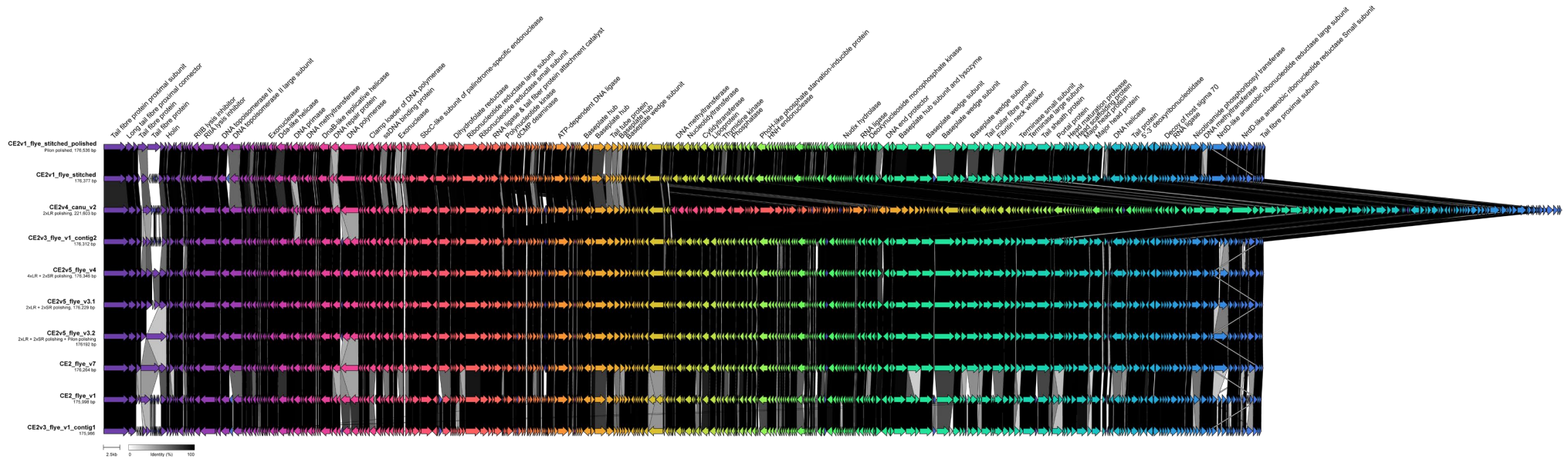
Figure 6.10. Gene cluster alignments for phage Fifoon (lab code CE2) long-read-only and HYPPA assemblies. Arrows represent coding sequences and are colour coded according to similarity. Pairwise comparisons are given as links (greyscale). Coding sequences without similarity are given in grey without links. Default display order is based on cluster similarity. Assembly used for further analyses was ce2v5_flye_v3.2. Figure generated with Clinker.

155

Multiple read mapping iterations were performed on different Fifoon assemblies using both the short- and long-reads (**Figure 6.11**) to check some of the more persistent errors. However, no conclusions could be drawn from any of the analyses as there were thousands of potentially incorrect basecalls or SNPs. For example, poor base support within reads at a given location in the assembly would be given as A at 49% and T at 51% (**Figure 6.11**). Based on these observations, it was hypothesised that Fifoon existed as a population of closely related genomes, rather than a clonal population that was observed for all other phage isolates in the collection, similar to the quasispecies concept known for RNA viruses with high mutation rates (Domingo and Perales, 2019; Lu *et al.*, 2020).

Without further curation, Fifoon was still designated a draft genome, and a conservative approach was taken when using the genome in further analyses.
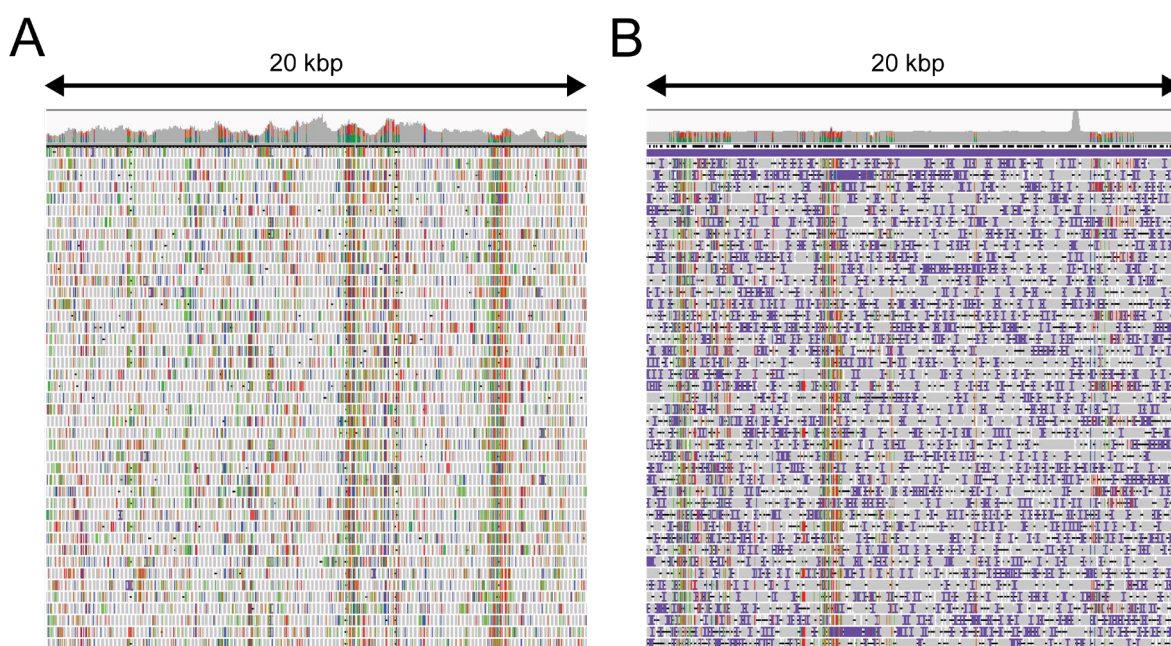
Figure 6.11. Read mapping of a 20 kbp section of the Fifoon draft assembly visualised in IGV. (A) Short-read data mapped back to the Fifoon draft assembly (ce2v5_flye_v3.2), performed with Bowtie 2. **(B)** Long-read data mapped back to the Fifoon draft assembly (ce2v5_flye_v3.2), performed with BWA-MEM. Read coverage is given (grey), with coloured lines representing poor base support for the base at that location (top panel). Stacking of individual reads for a particular location are given, with colours representative of the base called at that location. Purple indicates insertions. Base colours: green, A; red, T; blue, C; and orange-brown, G.

Resolving the genome of phage SteelHaze

Incorrect barcoding for phage SteelHaze (lab code CE14) resulted in the genome being found in the contigs of all other phages following Flye assembly. Three SteelHaze assemblies were extracted for comparison from three other randomly-selected phages: Suilven (CE6, 47,672 bp), Blaven (CE7, 47,669 bp), and DeeWhy (CE34, 47,677 bp), as well as its own assembly (47,667 bp, referred to as ce14_flye1_contig8) (**Table 6.4**). All contigs within the Suilven, Blaven, and DeeWhy assemblies (except the contig containing

the SteelHaze genome) were found to be bacterial contamination, and they could not be assembled with the HYPPA workflow, as described above. Additionally, 195 contigs of the assembly from which ce14_flye1_contig8 assembly was extracted were also bacterial contaminants. Following correct barcoding, a second long-read-only assembly for SteelHaze was generated with Flye (47,655 bp, referred to as ce14_flye2_contig5) (**Table 6.4**). The original short-read-only assembly was 47,756 bp (**Table 6.4**).

A total of 11 assemblies for SteelHaze were subsequently analysed and curated to generate a complete and accurate genome for this phage.

The HYPPA workflow was applied to both long-read-only SteelHaze assemblies (before and after correct barcoding), but not those extracted from other phage assemblies (**Table 6.4**). A preliminary MSA was performed to compare the nucleotide sequences for each assembly, including pre-polishing. Next, the assemblies were annotated and the coding sequences compared (**Figure 6.12**).

Table 6.4. Assembly comparisons for phage SteelHaze (lab code CE14) following long-read-only assembly with Flye and various stages of the HYPPA workflow.

Phage name	No. contigs	Assembly stage	Size (bp)
ce14_shovill	1	SR-only	47,756
ce14_flye1_contig8	196[a]	LR-only	47,667
ce14_flye1_contig8_v2	1	LR-only + LR polishing	47,673
ce14_flye1_contig8_v3	1	LR-only + LR + SR polishing*	47,673
ce14_flye2_contig5	5[a]	LR-only	47,655
ce14_flye2_contig5_v2	1	LR-only + LR polishing	47,672
ce14_flye2_contig5_v3	1	LR-only + LR + SR polishing*	47,672
ce14_flye2_contig5_v4	1	LR-only + LR + SR polishing + MC	47,677
ce14_from_ce6	63[a]	LR-only	47,672
ce14_from_ce7	263[a]	LR-only	47,669
ce14_from_ce34	82[a]	LR-only	47,756

LR, long-read; SR, short-read; MC, manual curation [a], genome contained within a single contig; *, only 1 iteration of short-read polishing performed (Polypolish).

Obvious errors in coding regions were observed in the portal protein, DNAP, RZ-like spanins, tail fibre proteins, and several hypothetical proteins (**Figure 6.12**). Following MSA and read mapping to look at errors in detail, both ce14_flye1_contig8_v3 and ce14_flye2_contig5_v3 appeared to be the most complete, with only one bp difference in their size. However, there was an error in both assemblies, in different regions. For ce14_flye1_contig8_v3, the error was in a hypothetical protein, whereas for ce14_flye2_contig5_v3, the error was in a virion structural protein towards the end of the assembly. Following read mapping, both areas were determined to be genuine errors that persisted following assembly with the HYPPA workflow. Thus, ce14_flye2_contig5_v3 was manually curated to correct the virion structural protein error, and renamed as ce14_flye2_contig5_v4, which was eventually the final SteelHaze genome (**Table 6.2**).
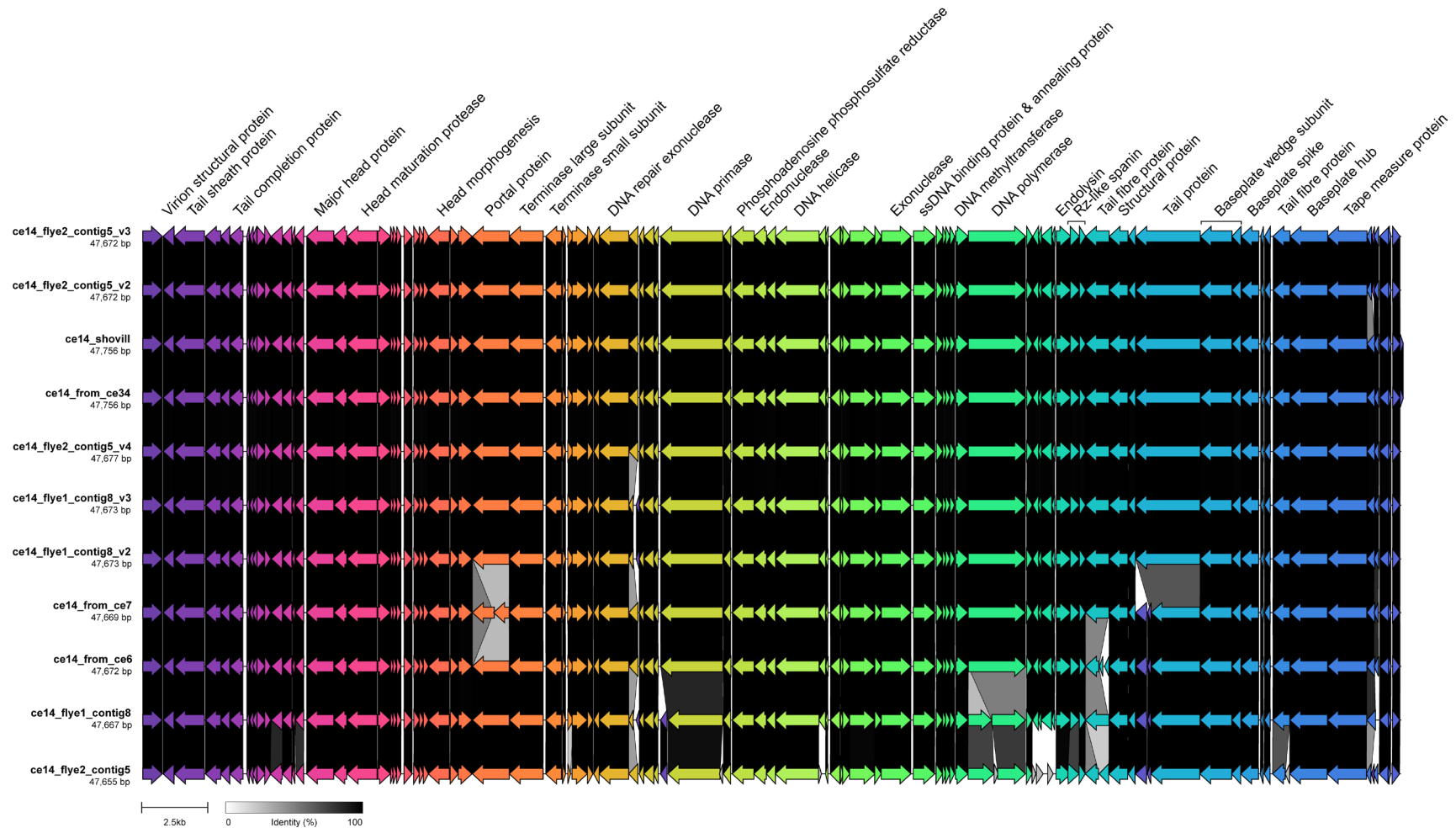
Figure 6.12. Gene cluster alignments based on protein coding sequences for phage SteelHaze (lab code CE14) short-read-only and long-read-only assemblies, plus assemblies at various stages of the HYPPA workflow. Arrows represent coding sequences and are colour coded according to similarity. Pairwise comparisons are given as links (greyscale). Coding sequences without similarity are given in grey without links. Default display order is based on cluster similarity. Final SteelHaze curated assembly used for further analyses was ce14_flye2_contig5_v4. Figure generated with Clinker.

Similar to the przondoviruses, the short-read-only assembly for SteelHaze (ce14_shovill) contained a 79 bp repeat region at the beginning and end of the genome. Moreover, the 79 bp repeat was found in one other long-read-only assembly (ce14_from_ce34), which appeared to be a repeat of a hypothetical protein (**Figure 6.12**). Erroneous repeat regions were identified within or close to the DTR in przondoviruses, regardless of where the genome was linearised, whereas for SteelHaze, erroneous repeat regions appeared to be due to the linearisation of the genome. However, SteelHaze represents a new genus and species of myovirus (see **section 6.3.4**), and it is unclear what type of packaging or genome organisation this phage has.

6.3 Phage genome characterisation and taxonomy

To provide a more detailed classification of the phages in the collection, a proteomic tree was constructed with whole-genome data against a reference database of Pseudomonadota-infecting phages using the final genomes draft, or otherwise, discussed above (**Figure 6.13**). Branches that went all the way to the centre represent genome clusters with no sequence similarity across their translated nucleotide sequences. Thus, the clustering of the phages within the proteomic tree demonstrates why phylogenetics alone cannot reliably infer relationships between phages, as even the most conserved genes share little to no amino acid sequence identity to generate an MSA. Whilst most of the phages within the collection were clustered together according to genera and/or family, this was not always the case. Indeed, SteelHaze demonstrated no similarity at all to the other myoviruses in the collection (**Figure 6.13**). SteelHaze does however, sit within a larger cluster of myoviruses, but which have not been classified at the family level. Similarly, phage Dereham was clustered further away from the other siphoviruses Iolaire, Fioreun, and DeeWhy (**Figure 6.13**). Reassuringly, all the phages preliminarily assigned to the genus *Przondovirus*, *Slopekvirus*, or *Jiaodavirus* clustered together within the family they were assigned to (**Figure 6.13**).

Viruses are particularly difficult to classify based on sequence similarity, whether nucleotide or protein, as they do not share universally conserved genes (Rohwer and Edwards, 2002), making phylogenetic classification difficult. Another approach to genomic classification of phages is through construction of a proteomic tree, allowing visualisation of phage diversity without the associated drawbacks of phylogenetic analyses (Rohwer and Edwards, 2002; Nishimura *et al.*, 2017).

Following assembly, preliminary taxonomic assignment was determined by BLASTn on all phages, identifying closest relatives in the NCBI GenBank database (as of February 2024), which showed varying levels of nucleotide similarity with reference phages (**Table 6.5**). Taxonomic assignment to a genus, where possible, was in accordance with the currently established ICTV genus demarcation of 70% nucleotide sequence similarity over the

genome length (Turner *et al.*, 2021). Where nucleotide sequence similarity fell below this threshold, it was determined that the phage belonged to a new genus within the same family (**Table 6.6**). Additionally, where phages showed ≥ 70% and < 95% sequence similarity over the genome length, they were assigned as different species within the same genus, whereas if they showed ≥ 95% nucleotide similarity over the genome length, there were assigned as different strains of the same species (**Table 6.6**).
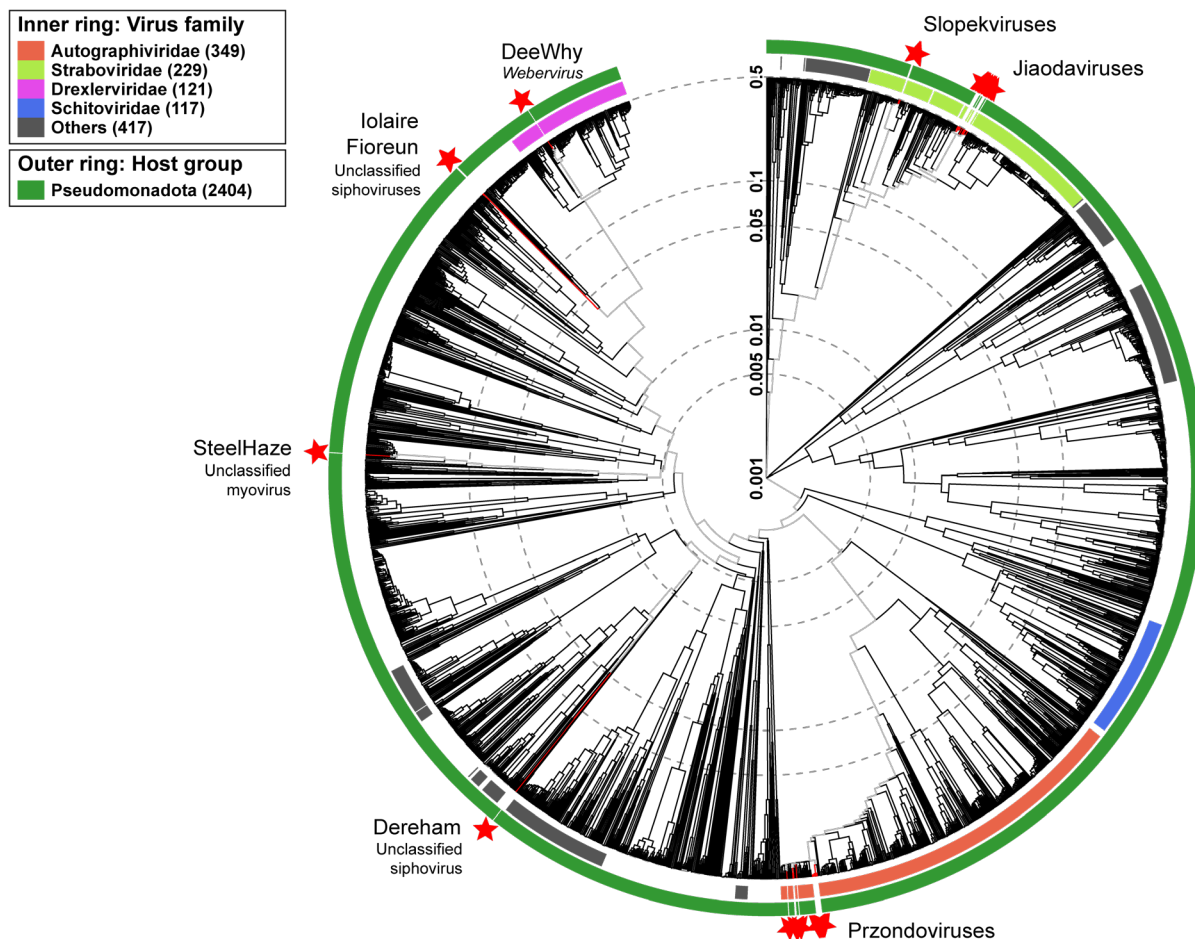
Figure 6.13. Proteomic phylogeny of phages in the collection against a database of dsDNA phages known to infect Pseudomonadota. Stars denote phages in the collection and are clustered according to proteomic similarity. Outer ring denotes host group (Pseudomonadota), inner ring denotes viral family. SteelHaze, Iolaire, Fioreun, and Dereham represent unclassified phages. Branch lengths represent pairwise dissimilarity scores on a logarithmic scale. Tree was constructed and visualised with ViPTree.

These data also further demonstrate the continued efforts to separate out the old morphology-based taxonomy (*Myoviridae*, *Podoviridae*, and *Siphoviridae*) to reflect phage genomics rather than their morphology (Turner *et al.*, 2021; Turner *et al.*, 2023). Currently, there are 48 families of tailed phages (class *Caudoviricetes*) that replace the original three morphology-based families *Myoviridae*, *Podoviridae*, *Siphoviridae*, which were abolished in 2022 (ICTV, 2023; Turner *et al.*, 2023). However, the VipTree database is not updated annually, and there are likely many phages within the VipTree database that will need to be re-assigned.

Table 6.5. Phages collection to date and data relating to the closest database relative.

Phage Name	Source	Isolation host	Genome size (bp)	Genome type	GC content (%)	No. of CDS	Closest database relative according to BLASTn		
							Name	Coverage (%)	ID (%)
Webervirus (*Drexlerviridae*)									
DeeWhy	Wastewater	K.va	50,630	Draft	50.93	87	*Klebsiella* phage P1	92.00	97.71
Slopekvirus (*Straboviridae*)									
Bolond	River water	K.mi	176,374	Complete	41.78	294	*Klebsiella* phage phi_KPN_S3	98.00	98.86
Fifoon	River water	K.pn	176,192	Draft	41.74	289	*Klebsiella* phage phi_KPN_S3	98.00	99.03
Jiaodavirus (*Straboviridae*)									
Assynt	Wastewater	K.pn	169,224	Draft	39.53	303	*Klebsiella* phage vB_KpnM_FRZ284	96.00	97.97
Torridon	Wastewater	K.pn	165,811	Draft	39.47	296	*Klebsiella* phage vB_KaeM_LilPanda	97.00	97.04
Suilven	Wastewater	K.pn	169,346	Draft	39.47	303	*Klebsiella* phage KP1	96.00	97.49
Blaven	River water	K.pn	167,086	Draft	39.50	290	*Klebsiella* phage JD18	96.00	98.21
Kandur	Wastewater	K.gr 1	164,832	Draft	39.51	284	*Klebsiella* phage KPV15	94.00	96.90
NejNjej	Wastewater	K.gr 1	165,293	Draft	39.52	284	*Klebsiella* phage KPV15	94.00	96.90
Pigeon	Wastewater	K.gr 1	164,671	Draft	39.56	288	*Klebsiella* phage KPV15	94.00	97.49
Sidders-16	Wastewater	K.gr 1	168,391	Draft	39.37	295	*Klebsiella* phage KPV15	93.00	96.05
Eggie	Wastewater	K.gr 1	164,529	Draft	39.53	286	*Klebsiella* phage Kpn6N	96.00	97.93
Przondovirus (*Autographiviridae*)									
Oda	River water	K.mi.	41,642	Complete	52.64	58	*Klebsiella* phage SH-KP152226	92.00	94.62
Toyotomi	Wastewater	K.mi.	41,268	Complete	52.64	55	*Klebsiella* phage SH-KP152226	92.00	94.71
Mera	Wastewater	K.mi.	41,400	Complete	52.58	56	*Klebsiella* phage SH-KP152226	92.00	94.34
Speegle	Wastewater	K.mi.	41,395	Complete	52.64	58	*Klebsiella* phage SH-KP152226	93.00	94.70
Cornelius	Wastewater	K.mi.	40,437	Complete	52.72	55	*Klebsiella* phage SH-KP152226	94.00	94.84

Tokugawa	Wastewater	K.mi.	41,414	Complete	52.64	56	*Klebsiella* phage SH-KP152226	92.00	94.70
Saitama	Wastewater	K.qp.	40,741	Complete	53.06	51	*Klebsiella* phage K11	96.00	95.71
Emom	Wastewater	K.gr 2	40,788	Complete	52.56	53	*Klebsiella* phage KP32	94.00	93.14
Amrap	Wastewater	K.gr 2	41,209	Complete	52.47	57	*Klebsiella* phage KPN3	85.00	95.05
Whistle	Wastewater	K.va.	40,735	Complete	52.40	54	*Klebsiella* phage IME264	94.00	94.78
Unclassified *Caudoviricetes*									
SteelHaze	River water	K.mi	47,677	Draft	49.43	89	*Klebsiella* phage BUCT_47333	73.00	91.08
Iolaire	Wastewater	K.qp	48,335	Complete	56.39	64	*Klebsiella* phage EKq1	99.00	99.11
Fioreun	Wastewater	K.qp	48,347	Complete	56.39	65	*Klebsiella* phage P61_2	99.00	99.16
Dereham	Wastewater	K.va	47,376	Draft	57.03	76	*Klebsiella* phage vB_Kpn_K21lambda1	88.00	93.69

K.gr 1, *K. grimontii* P038I; K.gr 2, *K. grimontii* M59 22 8 KoN (formerly *K. oxytoca*); K.mi., *K. michiganensis* M7 21 2 #2 WT; K.qp., *K. quasipneumoniae* P057K W; K.va., *K. variicola* DSM15968. CDS, coding sequences. Phage database relative accessions are given in **Appendix 10.1**. Accessions for the przondoviruses are given in Elek *et al.* (2023).

Table 6.6. Taxonomic rank thresholds for phages.

Rank	Similarity thresholds
Family	> 10% shared gene content
Genus	≥ 70% to < 95% sequence similarity
Species	≥ 95% to < 98% sequence similarity
Strain	≥ 98% sequence similarity

Based on these data phages were grouped into families and comparative genomics performed according to each family. Of the 26 novel phages isolated, ten were assigned to the genus *Przondovirus* within the subfamily *Studiervirinae* and family *Autographiviridae* family. Two phages were assigned to the genus *Slopekvirus* within the family *Straboviridae*, whereas nine were assigned to the genus *Jiaodavirus* within the subfamily *Tevenvirinae* and family *Straboviridae*. One phage was assigned to the genus *Webervirus* within the family *Drexlerviridae*. Three phages were unclassified members of the class *Caudoviricetes*, with two phages, Iolaire and Fioreun showing 98.5% nucleotide sequence similarity to one another. Both Iolaire and Fioreun also demonstrated ~97-98% nucleotide sequence similarity to two other database relatives, which are yet to be assigned a genus and family. One phage, SteelHaze likely represented a new genus as no closest relative was found that met the 70% nucleotide similarity threshold.

All phage genomes were analysed, with those in the genus *Przondovirus* being performed in more detail than the others. No toxin, virulence factor, or antibiotic resistance genes were found in any of the 26 phage genomes so far, and no integrases were found, suggesting that all phages were virulent. However, some phage genomes had a large number of hypothetical proteins encoded, which could have safety implications in phage therapy. All relatives were publicly available phage genomes and downloaded from the NCBI database.

6.3.1 The family *Drexlerviridae*: T1-like phage genomes

The only phage belonging to the genus *Webervirus* in the family *Drexlerviridae* was DeeWhy, whose draft genome was 50,630 bp with a GC content of 50.93%, which is lower than the average *Klebsiella* GC content of ~55-57% (**Table 6.5**). The genome contained 87 coding sequences (CDS) which were encoded bidirectionally (i.e. found on both the forward and reverse strands). Interestingly, no DNA polymerase gene was found for DeeWhy, and almost 30% of the genome consisted of hypothetical proteins.

To confirm the placement of DeeWhy within the genus *Webervirus* in the family *Drexlerviridae*, a nucleotide-based intergenomic similarity analysis was conducted with a selection of reference phages (**Figure 6.14**). Included in the analysis were relatives within the *Drexlerviridae* family, as well as unclassified siphoviruses Iolaire and Fioreun (see **section 6.3.5**) and a selection of their reference phages, as they demonstrated some proteomic similarities (i.e. branch lengths indicated some amino acid similarity).
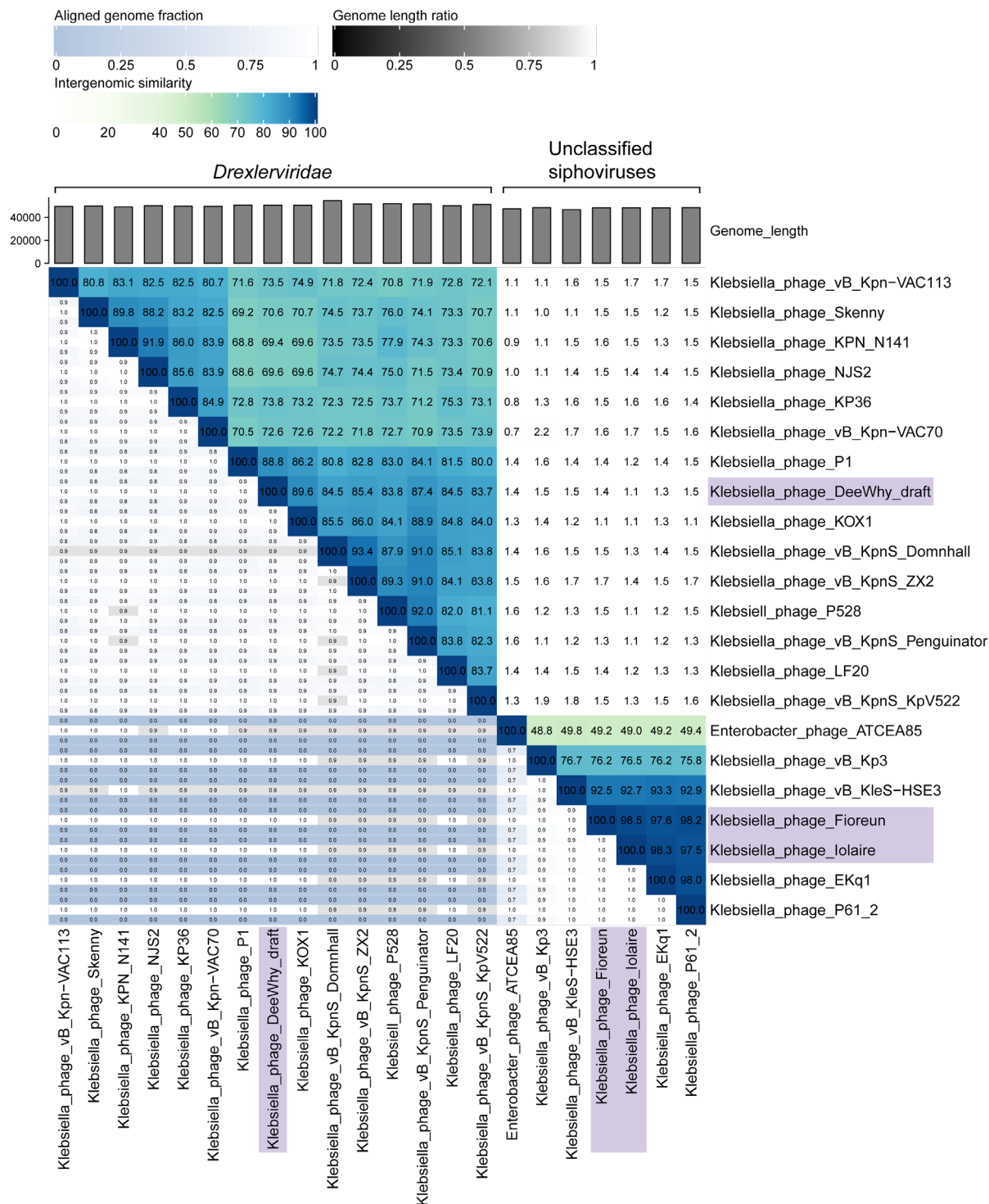
Figure 6.14. Comparative genomic analysis of phage DeeWhy and related weberviruses in the family *Drexlerviridae*, and phages Iolaire and Fioreun, and a selection of related unclassified siphoviruses. A nucleotide-based intergenomic similarities heatmap given as percentages (right half, blue–green heatmap). The aligned genome fraction for each phage pair in the row and column is represented by the top and bottom values, respectively (left, blue-grey scale). The genome length ratio for each phage pair is represented by the middle value (grey scale). Phages in the collection are highlighted (purple boxes). Heatmap generated with VIRIDIC.

The weberviruses clustered closely together within two distinct clades, where intra-clade similarity was within ~80-90%, and inter-clade similarity was within ~68-77% across the genome length (**Figure 6.14**). Based on the current ICTV genus threshold of 70% sequence

similarity across the genome length, it is possible that some database relatives should be re-assigned to different genera. However, complete and accurate genomes would be required for correct taxonomic reclassification, especially since some phages are at the upper end of the threshold. As expected, DeeWhy clustered with its closest relative (determined by BLASTn), *Klebsiella* phage P1, demonstrating a sequence similarity of 88.8%, but was more similar to *Klebsiella* phage KOX1 at 89.6% across the genome length. Together, these data suggest that DeeWhy should be assigned as the first exemplar of a new species of the same genus *Webervirus*, since the nucleotide similarity to the closest relative was below the same species threshold (≥ 95%).

The genomes of phages within the family *Drexlerviridae* are terminally redundant and circularly permuted, where the genome is packaged by the headful packaging mechanism. No transfer RNAs (tRNAs) were found in the genome of DeeWhy.

6.3.2 The family *Straboviridae*: T4-like phage genomes

The *Straboviridae* phages in the collection were preliminarily assigned to two different genera, *Slopekvirus* (two phages) and *Jiaodavirus* (nine phages) (**Table 6.2**). The genus *Jiaodavirus* is in the subfamily *Tevenvirinae*, which includes the model phage of the family *Straboviridae*, Enterobacteria phage T4, whereas the genus *Slopekvirus* does not have a subfamily assignment.

The slopekviruses had the larger genome size at ~176,000 bp and slightly higher GC content of ~41.7%, and the jiaodaviruses had genomes at 164,529-169,346 bp and a GC content of 39.37-39.56% (**Table 6.5**). For both genera, GC content was considerably lower than the average *Klebsiella* GC content of ~55-57%. The number of CDS reflected their large genome size, ranging from 289 to 294 for the slopekviruses and 284 to 303 for the jiaodaviruses. Interestingly, some jiaodaviruses encoded more proteins than the slopekviruses, despite being ~7,000-10,000 bp shorter than the latter. The CDS were encoded bidirectionally, and many were densely packed and/or overlapping. Additionally, some *Straboviridae* encoded multiple tRNAs. Bolond and Fifoon each encoded two tRNAs; all jiaodaviruses encoded 16 tRNAs each, except Torridon which encoded 15.

The genomic relationship between the two slopekviruses and nine jiaodaviruses in the collection and a selection of reference phages in the family *Straboviridae* was explored by conducting a nucleotide-based intergenomic similarity analysis (**Figure 6.15**). Included within the analysis were relatives of both genera (*Slopekvirus* and *Jiaodavirus*), and three from different genera, including those within the subfamily *Tevenvirinae* (*Karamvirus* and *Tequatrovirus*) and one without an assigned subfamily (*Pseudotevenvirus*). As expected, the slopekviruses and jiaodaviruses clustered closely together within their respective genera, where similarity within each genus was ~90-96%, and ~88-99%, respectively.

Phages Bolond and Fifoon were therefore correctly assigned to the genus *Slopekvirus*, and the jiaodaviruses in the collection were also correctly assigned to the genus *Jiaodavirus*.
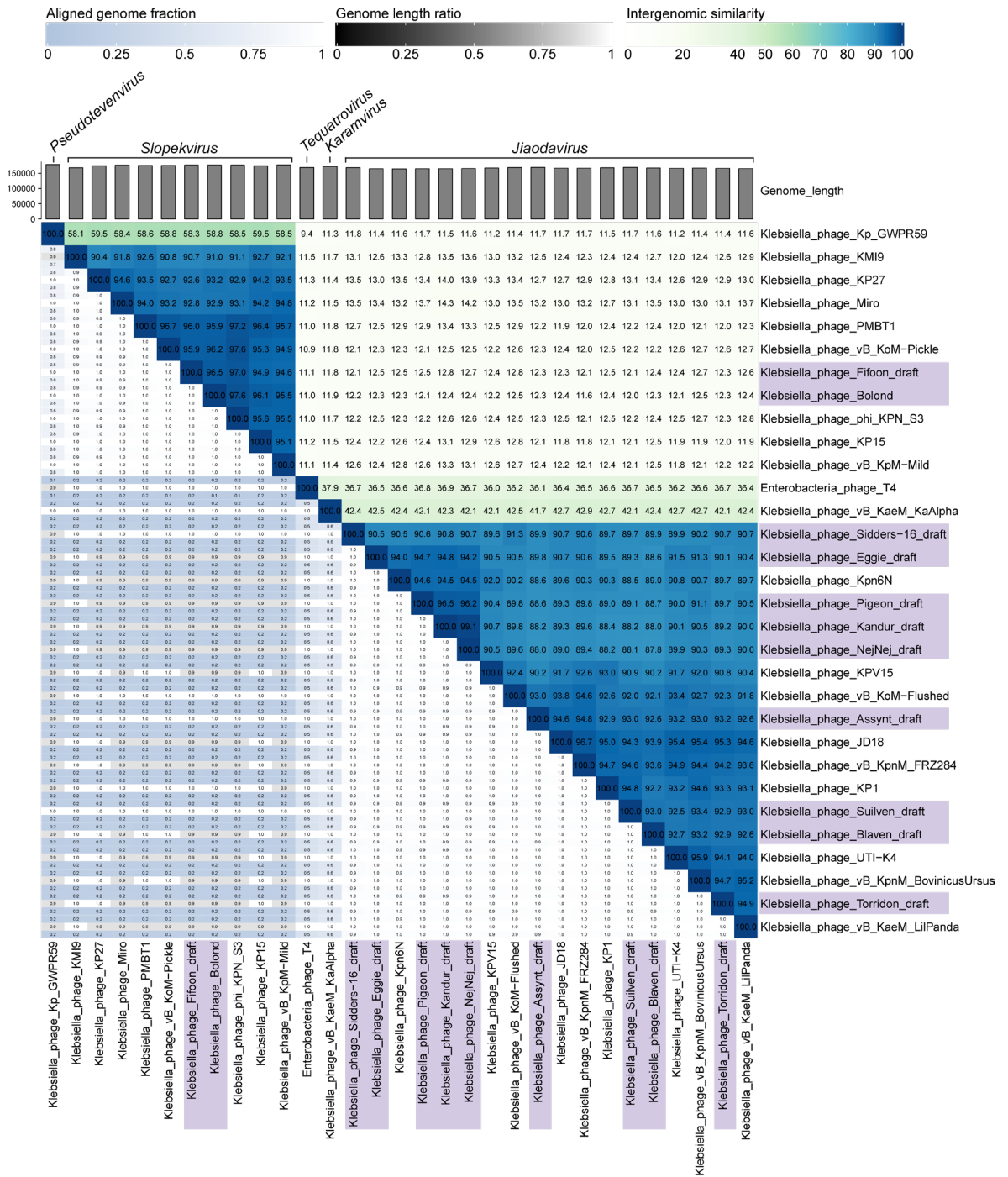
Figure 6.15. Comparative genomic analysis of *Straboviridae* phages in the collection and a selection of related phages spanning five genera. A nucleotide-based intergenomic similarities heatmap given as percentages (right half, blue–green heatmap). The aligned genome fraction for each phage pair in the row and column is represented by the top and bottom values, respectively (left, blue-grey scale). The genome length ratio for each phage pair is represented by the middle value (grey scale). Phages in the collection are highlighted (purple boxes). Heatmap generated with VIRIDIC.

Phages Bolond and Fifoon were isolated from the same river water sample using two different *Klebsiella* species (**Appendix 10.3**), but both could infect both hosts, and were initially suspected to be the same phage. However, they demonstrated 96.5% nucleotide

166

similarity across their genome length, suggesting that they could be two distinct strains within the same species. Moreover, both Bolond and Fifoon were ≥ 95% similar to *Klebsiella* phages PMBT1, vB_KoM-Pickle, and phi_KPN_S3, suggesting that all should be assigned to the *Klebsiella* virus PMBT1 species, which has not yet officially been defined. However, two phages, KP15 and vB_KpM-Mild met the species sequence similarity threshold for some of the phages listed above, but not others. A caveat to this is that the analyses have been performed with a draft assembly of Fifoon, and the genomes for reference phages could be incomplete. This further demonstrates the importance of accurate and complete genomes when assigning phages, particularly those close to the threshold for a specific taxonomic rank.

Phages Assynt, Torridon, and Suilven were all isolated from the same wastewater plant, whereas Blaven was isolated from a river water sample, and the remaining jiaodaviruses were isolated from a different wastewater plant (**Appendix 10.3**). Phages Assynt, Torridon, Suilven, and Blaven were more divergent than the other jiaodaviruses in the collection, sharing sequence similarity of ~88-90% across the genome length. As expected, they were more closely related to one another, having been isolated from different processing points of the same wastewater treatment plant, demonstrating ~92-93% sequence similarity across the genome length. Whilst they did cluster with their closest database relatives, none were within the 95% species demarcation. Based on these data, Assynt, Torridon, Suilven, and Blaven should be assigned separate new species within the genus *Jiaodavirus*.

Phages Kandur and NejNej demonstrated the highest sequence similarity across the genome length at 99.1%, having been isolated from the same wastewater treatment plant (**Appendix 10.3**). Pigeon also clustered closely to Kandur and NejNej, demonstrating ~96% sequence similarity, but was isolated from a different wastewater treatment plant (**Appendix 10.3**). Sidders-16 was isolated from the same wastewater treatment plant as Pigeon, but was slightly more divergent than all the other jiaodaviruses, at ~89-91% sequence similarity (**Appendix 10.3**). Similarly, Eggie demonstrated ~89-91% sequence similarity to other jiaodaviruses in the collection, but was isolated from a different wastewater treatment plant to the rest (**Appendix 10.3**). All were isolated against the same *Klebsiella* isolate. None of the jiaodaviruses in the collection met the species threshold with any of their closest database relatives. These data suggest that Kandur, NejNej and Pigeon should be assigned same new species within the genus *Jiaodavirus*, with Kandur and NejNej assigned to the same strain. Sidders-16 and Eggie should both be assigned separate new species within the genus *Jiaodavirus*. However, all taxonomic assignments for the jiaodaviruses in the collection are based on draft genomes, and could be subject to change if/when accurate and complete genomes are generated.

These data demonstrate the vast diversity across the nucleotide sequence for phages within the same family *Straboviridae*, with slopekviruses and jiaodaviruses sharing only ~12-14%

nucleotide similarity across the genome length. Moreover, the slopekviruses shared ~58% nucleotide similarity with *Klebsiella* phage Kp_GWPR59 (*Pseudotevenvirus*), and only ~11% nucleotide similarity with phages in the subfamily *Tevenvirinae*, but representative of different genera to the jiaodaviruses; *Escherichia* phage T4 (*Tequatrovirus*) and *Klebsiella* phage vB_KaeM_KaAlpha (*Karamvirus*). The opposite was true for the jiaodaviruses, which shared ~36-42% sequence similarity with phages in the same subfamily (*Tevenvirinae*) but different genera (*Tequatrovirus*, *Karamvirus*).

The genome of phage Fifoon could not be assembled using the short-read data alone, and even the long-read, traditional hybrid, and HYPPA assemblies proved to be problematic. It was hypothesised that this particular phage when replicating generates a population of highly genomically similar, but not identical phages, possibly caused by a mutation in the DNAP to generate a hypermutator phenotype that results in a quasispecies. To explore this further, the DNAP sequence of Fifoon was analysed (**Figure 6.16**).

First, the DNAP nucleotide sequence for the draft Fifoon assembly and the complete Bolond assembly were compared, revealing 35 mismatches (**Figure 6.16A**). Due to the conservative nature of the genetic code, only three mismatches resulted in differences in the amino acid sequence between the two phages. However, a single G present in the Fifoon sequence was not present in the Bolond sequence, resulting in a frameshift error and subsequently one hypothetical and two DNAP annotations in the former. Mapping the reads back to the DNAP sequence of Fifoon revealed an insertion error as the site of the single G, confirming this was indeed a genuine error (**Figure 6.16B**). However, the authenticity of the remaining mismatches is yet to be fully elucidated. There were several differences in the CDS between the different Fifoon assemblies, and six of the ten assemblies analysed contained the erroneous DNAP (see **Figure 6.10** in **section 6.2.9**). Indeed, gene clustering of the DNAP coding sequence revealed considerable differences between Fifoon when compared to that of Bolond, another Fifoon DNAP (extracted from CE2v1_flye_stitched_polished), and the DNAP of five other *Straboviridae* reference phages (**Figure 6.16C**). The nucleotide sequence of the DNAP extracted from the Fifoon draft assembly was less divergent than the CDS, being slightly lower in sequence similarity (~98.6%) than the DNAP extracted from the other slopekviruses (**Figure 6.16D**).

Besides the insertional error, the remaining mismatches between the DNAP of Fifoon and that of Bolond could be genuine single nucleotide polymorphisms that result in a more mutagenic DNAP, which could lead to a population of Fifoon genomes that differ in a few to many nucleotides across the genome length i.e. a quasispecies. This could explain why the genome of Fifoon has been difficult to resolve, where multiple assemblies have produced multiple differences. Due to time constraints however, the mutagenic DNAP hypothesis could not be fully explored, and no conclusions determined.
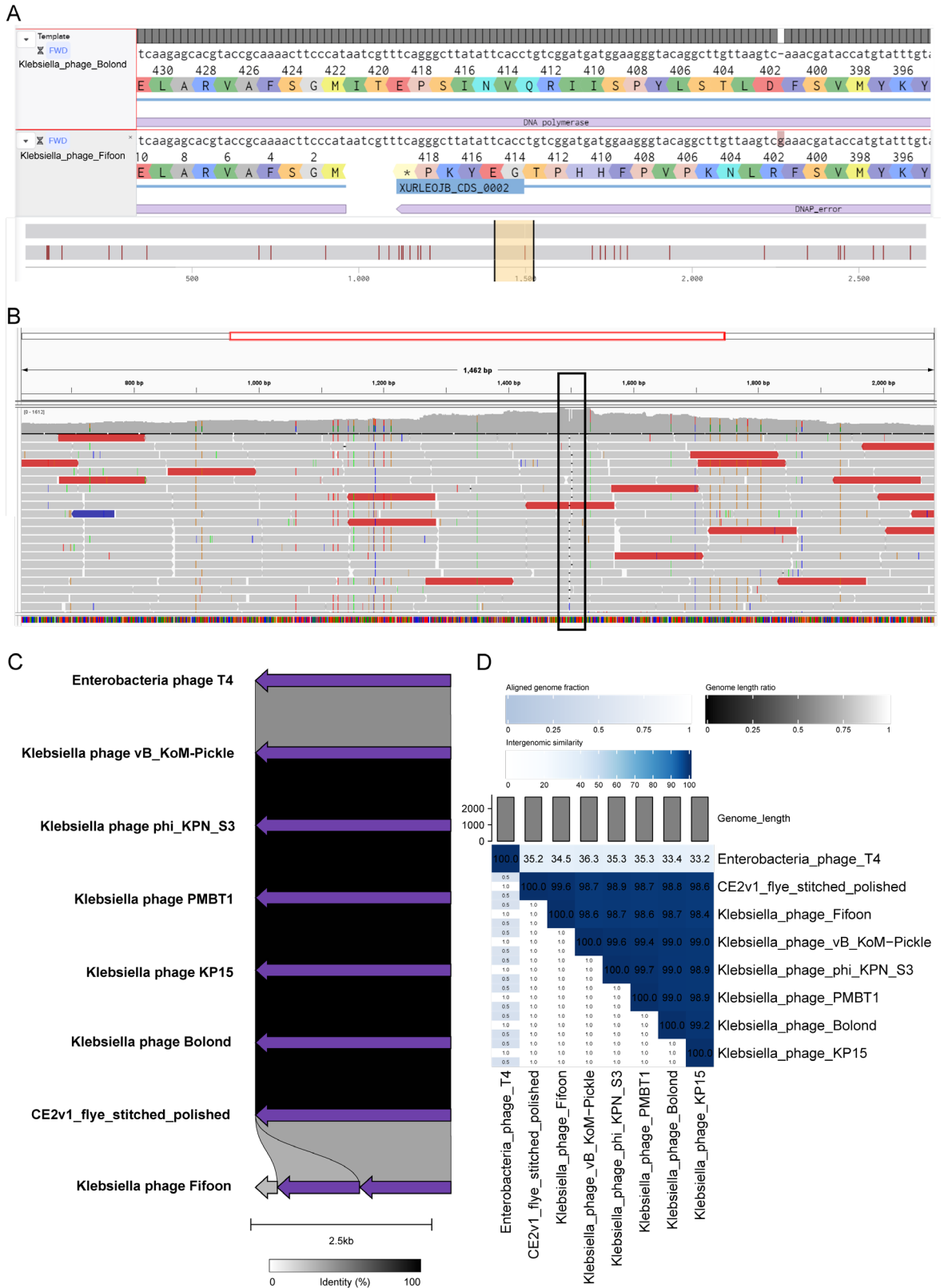
Figure 6.16. Divergence of the DNAP extracted from the draft Fifoon assembly compared to its closest relative in the collection Bolond and a selection of reference phages. (A) Basic alignment of the DNAP nucleotide sequences performed with Benchling and extracted from the complete Bolond assembly (top panel) and the Fifoon draft assembly (middle panel) showing the single G single-nucleotide error and subsequent frameshift error. 35 mismatches were identified across the genome length of Fifoon when compared to Bolond (red lines, bottom panel). **(B)** Read mapping of the short-read data back to the DNAP of Fifoon and the insertional error (black box) performed with IGV. **(C)** Gene clustering generated with Clinker of the DNAP extracted from the Fifoon draft assembly

(*Klebsiella* phage Fifoon) and another Fifoon assembly without the error (CE2v1_flye_stitched_polished), as well as Bolond, five others in the family Straboviridae, including four other slopekviruses, and Enterobacteria phage T4 (Tequatrovirus). **(D)** A VIRIDIC heatmap of the nucleotide-based intergenomic similarities of the DNAP extracted from the aforementioned phages given as percentages (right half, blue heatmap).

Similar to the phages in the family *Drexlerviridae*, phages in the family *Straboviridae* have terminally redundant and circularly permuted genomes, with headful packaging.

Many T4-like viruses encode a large number of tRNAs. A large number of tRNAs were found within the jiaodavirus genomes, but not as many were identified in the slopekvirus genomes. tRNA is involved in translation of mRNA to protein, and many phages utilise the host tRNAs as they do for all other cellular processes in order to replicate (Bailly-Bechet *et al.*, 2007; Yang *et al.*, 2021). However, as host takeover continues, so does degradation of cellular machinery (Yang *et al.*, 2021). Thus, there is a selective pressure for translation to continue as the lytic cycle continues (Yang *et al.*, 2021). Moreover, phage-encoded tRNAs exhibit codon optimisation, similarly to their bacterial hosts, which is where particular codons are used more often than others and demonstrate bias towards translating their own proteins (Rocha, 2004; Bailly-Bechet *et al.*, 2007; Yang *et al.*, 2021). Additionally, phage-encoded tRNAs are thought to allow for large differences in GC content between the phage and the host (Bailly-Bechet *et al.*, 2007), and more recently, are involved in counteracting host-mediated anti-phage defence systems that degrade phage-encoded tRNAs (van den Berg *et al.*, 2023). Indeed, the lowest GC content was observed for the jiaodaviruses, whereas the slopekviruses had a slightly higher GC content (**Table 6.5**).

6.3.3 The family *Autographiviridae*: T7-like phage genomes

Ten of the 26 phages in the collection were preliminarily assigned to the genus *Przondovirus* in the subfamily *Studiervirinae* and family *Autographiviridae*, for which complete and accurate genomes had been generated for all. The family also contains the exemplar *Escherichia* phage T7. The przondoviruses represented the smallest phage genomes of the collection at 40,336 to 41,720 bp (**Table 6.5**). They also had a higher GC content than many of the other phages in the collection, at ~55.46-57.59%, and no tRNAs were found in their genomes. The number of CDS also reflected their smaller genome size, which varied from 51 to 58. Almost all CDS were unidirectional, that is, they were found in a left to right direction on the forward strand, with between one and four hypothetical proteins identified on the opposite strand. However, due to the highly conserved genome organisation where all genes are unidirectional, further laboratory and/or bioinformatic analyses would be required to determine the authenticity of the proteins encoded on the opposite strand.

All early, middle, and late genes associated with viral host takeover, DNA replication, and virion assembly, respectively were found in the complete przondovirus genomes from the

collection (**Figure 6.17**). A good marker for the start of the genome was previously identified as S-adenosyl-L-methionine hydrolase, and was one of the first annotated proteins. However, this protein was not found in *Enterobacteria* phage K30, but the functionally-related Ocr protein was: both proteins are involved in counteracting against host-mediated RM systems (Guttman *et al.*, 2005; Molineux, 2006; Severinova and Severinov, 2006; Roberts *et al.*, 2012; Guo *et al.*, 2021). Importantly, all complete genomes contained the DNA-directed RNA polymerase (RNAP) (**Figure 6.17**), which reduces the reliance on host-encoded RNA polymerases for middle and late gene expression (Choi, 2012). Studies in T7 have shown that only a very small proportion of phage DNA enters the cytoplasm following infection, with the rest being pulled into the cell by the transcription activity of the phage-encoded RNAP (Savalia *et al.*, 2010).

Middle genes found in the complete genomes were typical for DNA replication, and included genes for inhibiting host-mediated DNA replication and expression, thereby switching to phage-mediated DNA replication and expression (**Figure 6.17**). Late genes were also typical for T7-like phages, consisting of all the genes necessary for virion assembly (capsid and tail-associated genes) and host lysis (holins, Rz-like genes) (**Figure 6.17**).

Przondoviruses in the collection that were isolated against the same *Klebsiella* strains shared considerable protein sequence similarity across their entire genome. For example, phages Oda, Toyotomi, Mera, Speegle, Cornelius, and Tokugawa (isolated against the same *K. michiganensis* strain), and Emom and Amrap (isolated against the same *K. grimontii* strain), and is highlighted by the high homology across their tail fibre proteins. These data are corroborated by nucleotide-based intergenomic similarities (see **Figure 6.18**).

The tail fibre and tail spike proteins are major determinants for host range, so high sequence similarity between the tail fibre proteins was expected between phages isolated against the same *Klebsiella* host (**Figure 6.17**). However, in some przondoviruses, a hypothetical protein followed the first tail fibre protein, and protein BLAST (BLASTp) revealed high homology to tail spike proteins of other *Klebsiella* przondoviruses in the database.

Phage Toyotomi however, was able to infect two hosts in our collection, whereas the remaining closely-related phages from the collection were not. This is intriguing since the high similarity in the tail fibre protein would suggest similar narrow host range capabilities for this subset of przondoviruses in the collection. Contrastingly, despite infecting different host species, Mera (and by extension the other closely related przondoviruses in the collection) and its closest database relative SH-KP152226 shared high sequence similarity across the entire genome, including at the tail fibre location. Both Saitama and Whistle also shared considerable similarity in their tail proteins with their closest relatives, but Emom (and by extension Amrap) did not, which is not explained by differences in host range between Emom and Amrap and *Klebsiella* phage KP32.
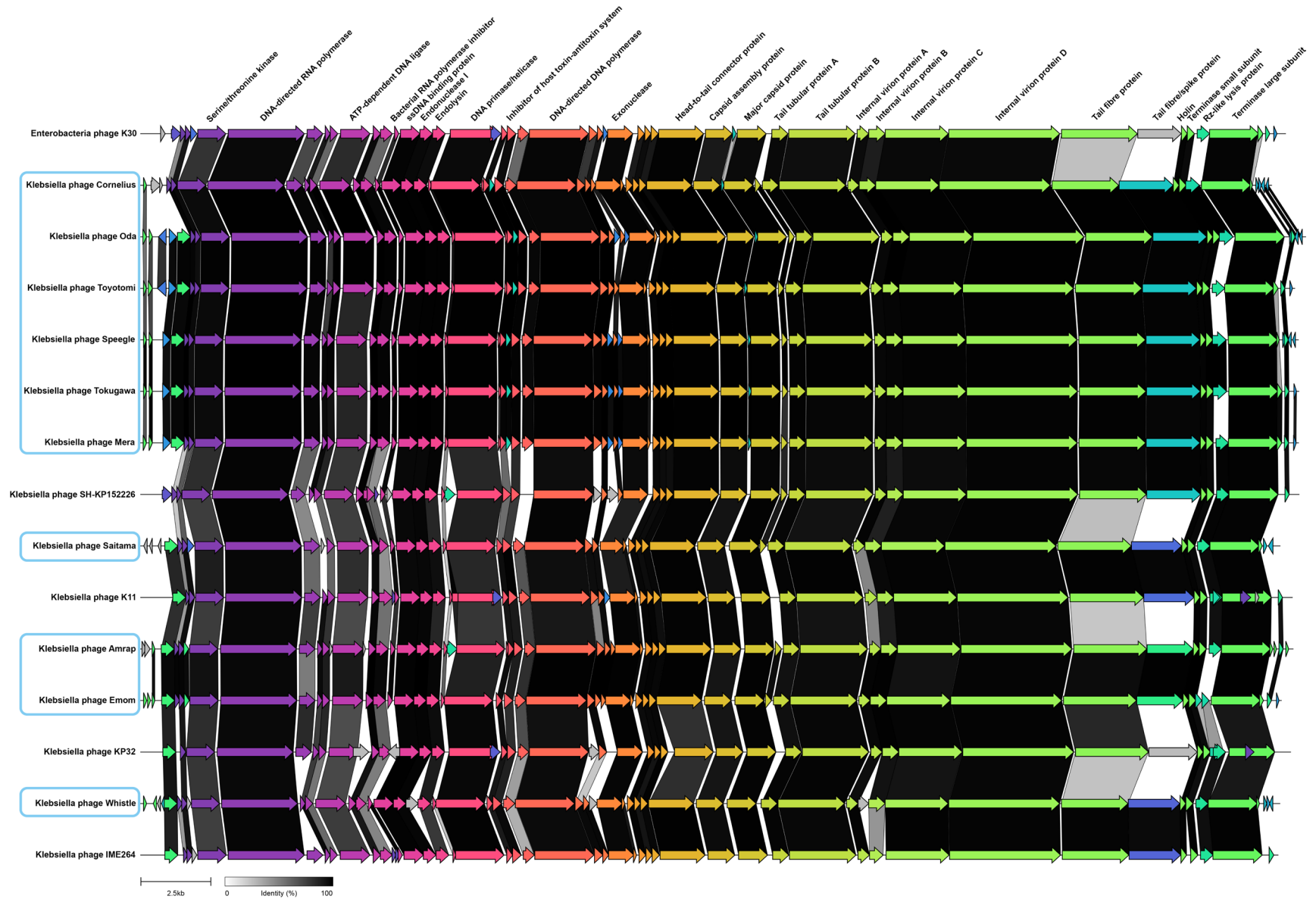
Figure 6.17. Gene cluster alignments based on protein coding sequences of accurate and complete genomes for the przondoviruses in the collection and a selection of related phages. Arrows represent coding sequences and are colour coded according to similarity. Pairwise comparisons are given as links (greyscale). Coding sequences without similarity are given in grey without links. DTRs were not annotated. Przondoviruses in the collection are highlighted (blue boxes). Figure generated with Clinker.

The most striking differences in tail proteins were between przondoviruses in the collection and more divergent (i.e. not closest database relatives) reference phages in the same genus. Only partial sequence similarity was observed between Saitama and SH-KP152226 in the first tail fibre for example, and no sequence similarity was observed between the two in the second tail fibre protein. Again, this phenomenon was not explained by the differences in host range, which would be expected in this case, but not in others that did show high sequence similarity across both tail fibre/spike proteins.

Another major difference observed was between the Rz-like proteins of closely related phages from the collection. For example, no sequence similarity was observed in this protein between Oda, Toyotomi, and Speegle, or Tokugawa and Mera, but was between Tokugawa and Speegle, and also Cornelius and Oda. Similarly to tail fibre proteins, Rz-like proteins can be highly diverse, and are involved in lysing the bacterial cell by disruption of both the inner and outer membranes (Summer *et al*., 2007; Berry *et al*., 2008; Kongari *et al*., 2018). They can be part of a single- or two-component system, where the latter has multiple subtypes, whose evolutionary lineage appear to be distinct from one another (Kongari *et al*., 2018). A possible explanation for the divergence in sequence similarity could be due to differences in membrane composition among the four *Klebsiella* spp. used to isolate the przondoviruses in the collection. Another possible explanation could be distinct evolutionary origins of the proteins themselves.

The ten przondoviruses were isolated from multiple samples across various locations and against four different *Klebsiella* isolates (**Appendix 10.3**). To understand their genomic relationship, a nucleotide-based intergenomic similarity analysis was conducted (**Figure 6.18**). The analysis included relatives within the genus *Przondovirus*, relatives within the same subfamily (*Studiervirinae*) but across four different genera (*Apdecimavirus*, *Berlinvirus*, *Teetrevirus*, *Teseptimavirus*).

Przondoviruses isolated against the same *Klebsiella michiganensis* strain clustered closely together, despite being isolated from different sources (**Figure 6.18**). For example, *Klebsiella* phage Oda (river water) clustered closely with phages isolated from the same wastewater plant: Toyotomi, Mera, Speegle, Cornelius, and Tokugawa. All were within ~98% similarity across their genome length, except Cornelius, which was within ~95-96% similarity across the genome length and whose genome was approximately 830 to 1200 bp shorter than the other cluster of przondoviruses. Whilst the phages did cluster with their

closest database relative *Klebsiella* phage SH-KP152226, it was below the same species threshold ≥ 95% (~87-88%). Based on these data, this collection of six phages were assigned as different strains of the same proposed new species *Przondovirus oda* pending ratification by the ICTV.
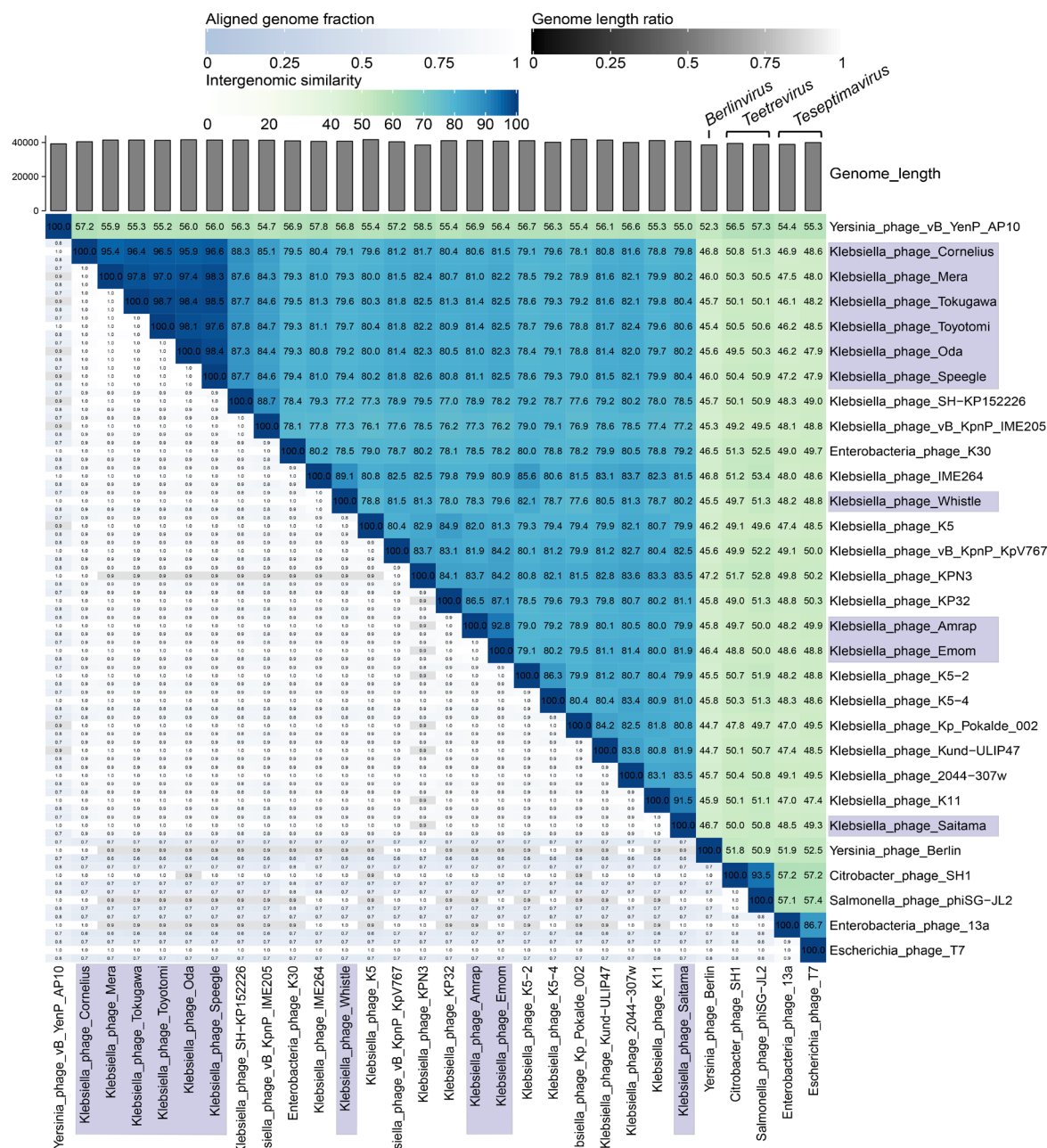
Figure 6.18. Comparative genomic analysis of przondoviruses in the collection and a selection of related phages within the subfamily *Studiervirinae.* A nucleotide-based intergenomic similarities heatmap given as percentages (right half, blue–green heatmap). The aligned genome fraction for each phage pair in the row and column is represented by the top and bottom values, respectively (left, blue-grey scale). The genome length ratio for each phage pair is represented by the middle value (grey scale). Phages in the collection are highlighted (purple boxes). Yersinia phage vB_YenP_AP10 is representative of the genus *Apdecimavirus*. Heatmap generated with VIRIDIC.

Klebsiella phages Saitama, Emom, Amrap, and Whistle did not cluster closely with the other przondoviruses in the collection, showing nucleotide similarities of ~79-82% (**Figure 6.18**). Emom and Amrap did however, cluster closely together and were within ~92% similarity, also having been isolated against the same *K. grimontii* strain but from two different wastewater plants (**Figure 6.18**). They also clustered with their closest database relative *Klebsiella* phage KP32, but again was below the same species threshold. Based on these data, *Klebsiella* phages Emom and Amrap were proposed to be assigned separate new species *Przondovirus emom* and *Przondovirus amrap*, respectively, pending ratification by the ICTV.

Whilst Saitama and Whistle clustered to their respective closest relatives *Klebsiella* phage K11 and *Klebsiella* phage IME264, respectively, they were also below the same species threshold of ≥ 95% (at 91.5% and 89.1%, respectively) (**Figure 6.18**). Thus, Saitama and Whistle were each assigned to a new proposed species *Przondovirus saitama* and *Przondovirus whistle*, respectively, also pending ratification.

A high degree of diversity was also observed across genera belonging to the same subfamily *Studiervirinae*, where inter-genera nucleotide similarity between *Apdecimavirus*, *Berlinvirus*, *Teetrevirus*, *Teseptimavirus* and przondoviruses in the collection was only ~45-57% (**Figure 6.18**).

All annotated genes are highly conserved across all *Przondovirus* genomes and several sequencing and/or assembly errors were identified among some przondovirus closest database relatives, resulting in potentially incomplete genomes (see **Figure 6.1** in **section 6.2.1**). Such incomplete genomes could skew the comparative genomics analyses, emphasising the importance of careful curation of phage assemblies, particularly when the genomes are close to taxonomic thresholds. Specifically, potential sequencing and/or assembly errors included i) missing right terminal repeats; ii) missing genes; iii) incomplete annotation; and iv) incorrect start site.

For example, *Klebsiella* phages KP32 (accession MN101227), KPN3 (accession MN101227), and IME264 (accession OL799328) were missing the right terminal repeat sequences, making their genomes shorter at 41,119 bp, 38,503 bp, and 40,671 bp, respectively. Phage KPN3 in particular, was ~3,000 bp shorter than all other przondoviruses with only 45 CDS. Missing genes included no annotated DNA-directed DNAP in the genome of KPN3, leaving a large area of the genome without any CDS and just a few hypothetical proteins, and a missing tail tubular protein A (see **Figure 6.1** in **section 6.2.1**). *Klebsiella* phage KMI1 (accession MN052874) contained a truncated DNA-directed RNAP and was missing several other proteins, including the S-adenosyl-L-methionine hydrolase and serine/threonine kinase genes (see **Figure 6.1** in **section 6.2.1**). *Klebsiella* phage SH-KP152226 was also missing the S-adenosyl-L-methionine hydrolase gene (see **Figure 6.1**

in **section 6.2.1**). KMI1 had an incorrect start site, with the first protein encountered in the assembly being the terminase large subunit (TerL) (see **Figure 6.1** in **section 6.2.1**).

Typically, phage phylogeny is determined using the protein sequence of a single conserved gene, often the TerL, or a set of core genes. However, phylogenetic analysis of the przondoviruses in the collection was performed using the protein sequence of the DNA-directed RNAP, since it is a hallmark of the family *Autographiviridae*. To further verify the taxonomic classification of the przondoviruses, the protein sequence of the DNA-directed RNAP was extracted from database relatives within the subfamily *Studiervirinae* (**Figure 6.19)**.
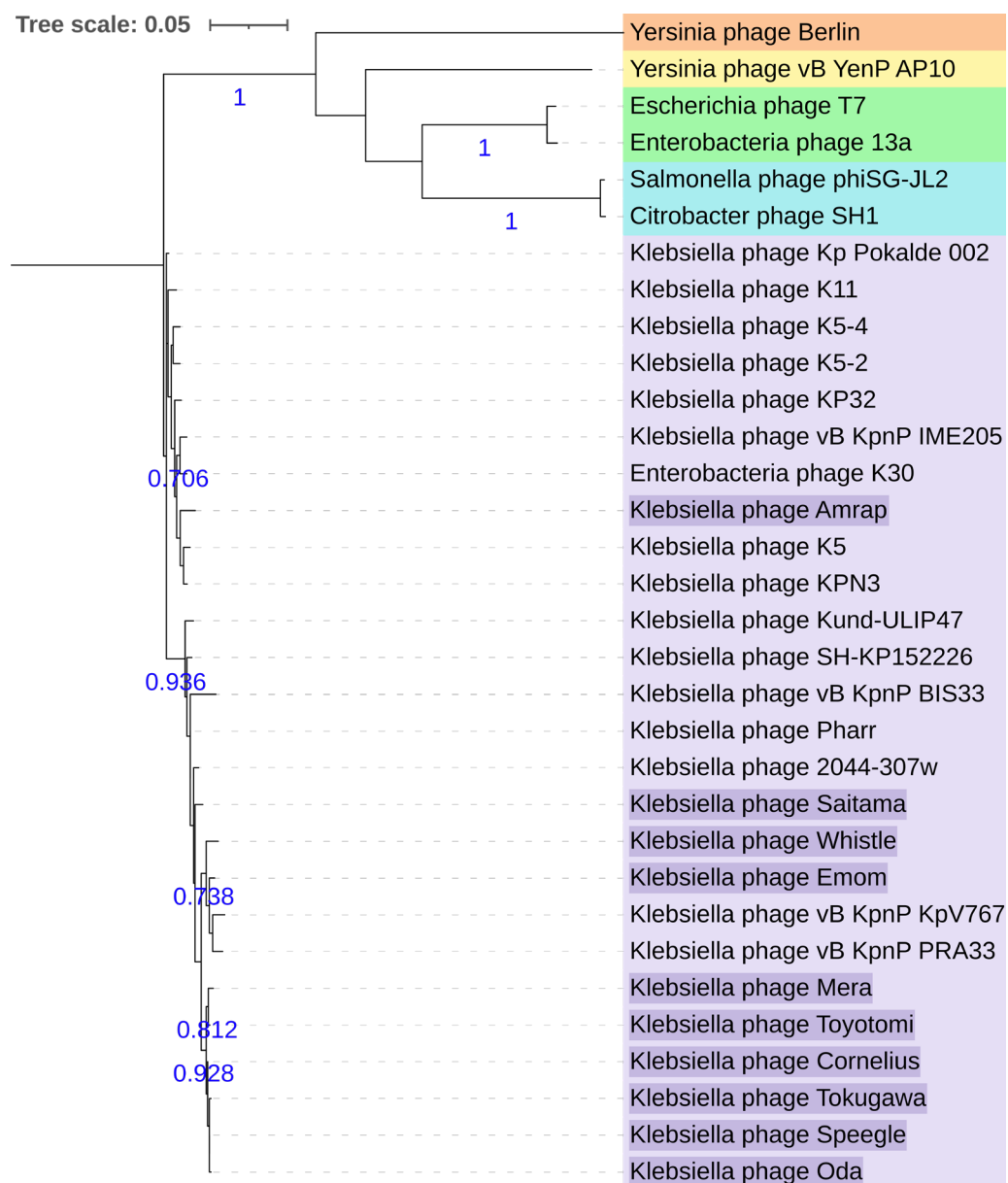
Figure 6.19. Maximum-likelihood phylogeny of the RNAP protein sequence of przondoviruses in the collection and a selection of related phages within the subfamily *Studiervirinae*. Przondoviruses in the collection are highlighted (dark purple) and clustered with other przondoviruses (light purple). Outgroups are represented by genera *Berlinvirus* (orange), *Apdecimavirus* (yellow), *Teseptimavirus* (green), and *Teetrevirus* (blue). Tree is midpoint rooted. Boostrap support values at ≥ 0.7 are given in

blue (500 replicates). MSA of the RNAP performed with MUSCLE in MEGA X and tree rendering performed with iTOL.

Przondovirus phylogeny confirmed the nucleotide comparative genomic analysis, where phages within the same genus clustered more closely together and a clear delineation of phages from different genera (**Figure 6.19**). *Klebsiella* phage KMI1 was excluded from the analyses due to the truncated RNAP.

6.3.4 The unclassified myoviruses

The only unclassified myovirus was SteelHaze, whose draft genome was 47,677 bp with a GC content of 49.43%, which is lower than the average *Klebsiella* GC content of ~55-57% (**Table 6.5**). The genome contained 89 CDS, which were bidirectional, of which 53 were designated hypothetical proteins, meaning that nearly 60% of the proteins were of unknown function.

An initial BLASTn (as of February 2021) of the genome against all phages in the nucleotide database found few close database relatives. Preliminary analyses found that SteelHaze could represent a new genus of phage. A new BLASTn was performed (as of January 2024), and more close database relatives were found having been submitted to the NCBI GenBank database recently. To explore the genome organisation and synteny of SteelHaze and its closest database relatives within the genera *Jedunavirus* and *Peatvirus*, a cluster gene alignment of the protein CDS was performed (**Figure 6.20**). Initially, all genomes had to be reorientated to the same starting sequence, which was difficult considering that they were considerably divergent (as demonstrated by the intergenomic relationship analyses discussed in **Figure 6.21**). The small terminase subunit (*terS*) was the easiest to find in all relatives based on nucleotide and amino acid sequence similarity. Thus, the genome of SteelHaze was reorientated to start with the *terS* gene (**Figure 6.12**).

Some early, middle, and late genes were found in SteelHaze (see **Figure 6.12** in **section 6.2.9** and **Figure 6.20**) and relatives, with some proteins being more highly conserved than others. For example, the portal protein, tail sheath, tail fibre, baseplate proteins, and the ssDNA binding and annealing protein all exhibited high protein sequence similarity. Some degree of sequence similarity was observed for capsid (head) proteins, some tail fibre proteins, an exonuclease, and DNA primase, among others. Interestingly, some phages showed high sequence similarity in their TerL, with no sequence similarity at all between others. Similarly, the tail protein among four relatives exhibited high sequence similarity (teal, **Figure 6.20**). However, two proteins in the same location as the tail protein for phages BUCT_47333 and JD001 exhibited high sequence similarity with one another, but were not similar to the tail proteins, having been annotated as a gluconolaconase (orange, **Figure 6.20**).
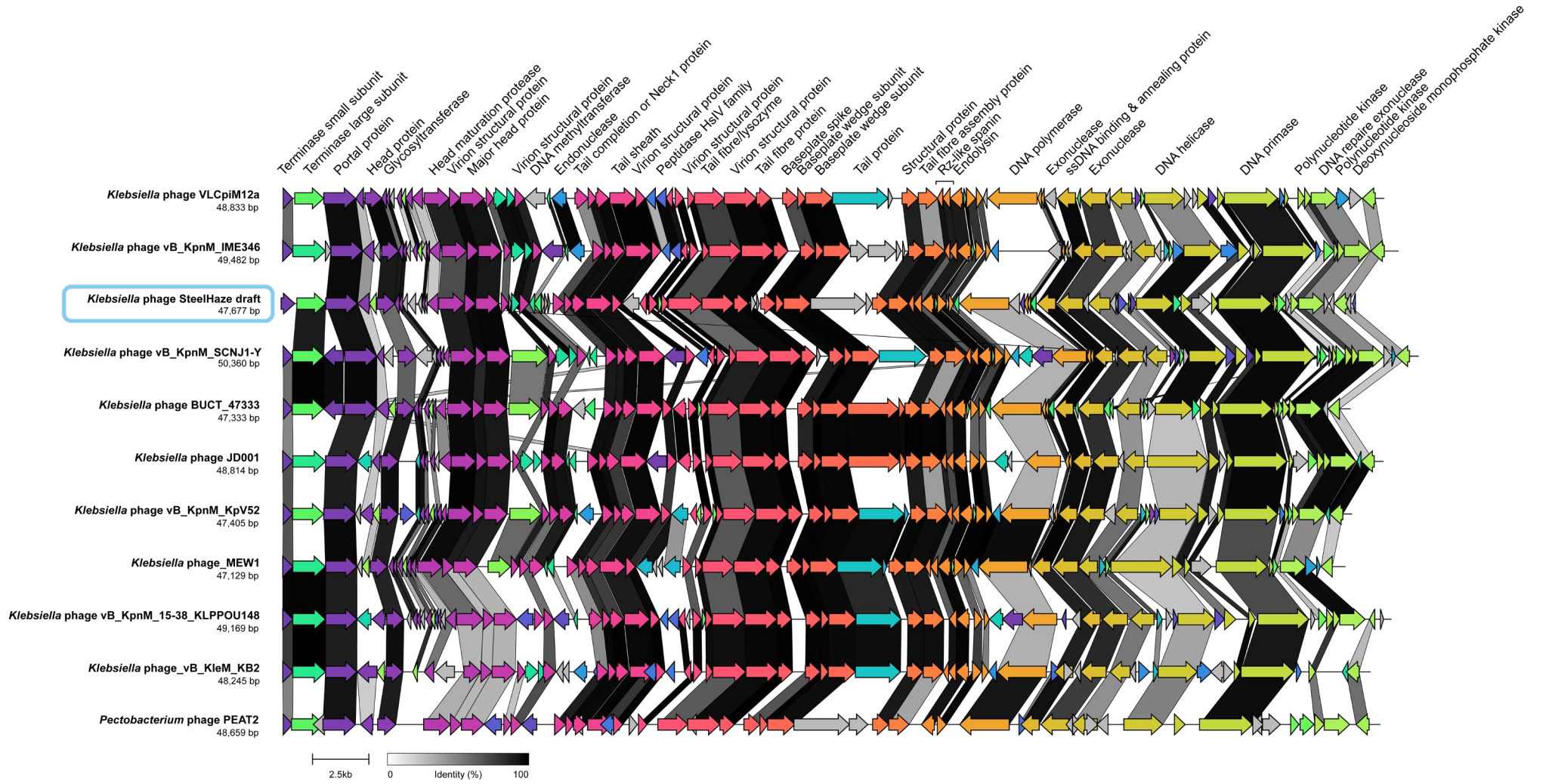
Figure 6.20. Gene cluster alignments based on protein coding sequences for phage SteelHaze and a selection of related unclassified myoviruses. Arrows represent coding sequences and are colour coded according to similarity. Pairwise comparisons are given as links (greyscale). Coding sequences without similarity are given in grey without links. Default display order is based on cluster similarity. Annotations are given for *Klebsiella* phage VLCpiM12a, meaning that some proteins present in other genomes have not been annotated. The SteelHaze genome is highlighted (blue box). Figure generated with Clinker.

Interestingly, the Rz-like spanins were conserved across almost all relatives (**Figure 6.20**), in contrast to the Rz-like spanins of the przondoviruses, despite the divergence of these unclassified myoviruses

Contrastingly, the DNAP was also conserved in all but one phage, *Klebsiella* phage vB_KpnM_IME346, but showed higher divergence in its sequence similarity for some phages, but not others (**Figure 6.20**). For example, *Klebsiella* phage vB_KpnM_KpV52 shared high sequence similarity with the DNAP of *Klebsiella* phage MEW1, as did *Klebsiella* phage vB_KleM_KB2 and *Pectobacterium* phage PEAT2. The remaining relatives shared < 50% sequence similarity in their DNAP with their neighbours and many varied in size. It is possible that the genomes have not been appropriately curated however, leading to errors in the DNAP and possibly many other proteins. It is also possible however, that the DNAP is divergent across the entire group.

Since taxonomy is now based on genomic data rather than morphology, there are a vast number of unclassified phage genomes within the database directly assigned to the class *Caudoviricetes* which contains all tailed phages (Turner *et al.*, 2023). These unclassified phages sometimes have a genus assigned but have no formal taxonomic classification at the family rank.

To explore the genomic relationship between SteelHaze and a selection of related unclassified myoviruses, a nucleotide-based intergenomic similarity analysis was performed with genomes downloaded from the NCBI database, none of which showed high similarity to one another across their genome length (**Figure 6.21**). None of the closest database relatives were assigned to a family, having been previously extracted from the now obsolete family *Myoviridae*. *Pectobacterium* phage PEAT2 belonged to the genus *Peatvirus*, and the rest belonged to the genus *Jedunavirus*, except vB_KpnM_SCNJ1-Y, whose taxonomy at the time of writing was unknown. Some of the relatives were within the 70% genus demarcation, whereas others were not. Phage SteelHaze demonstrated ~56-66% sequence similarity across the length of the genome with reference isolates, clustering most closely with its closest relative identified by BLASTn *Klebsiella* phage BUCT_47333 (66.8%). As the 70% intergenomic distance genus threshold was not met however, these data suggest that it should be assigned to a new genus within the same family as the related

phages. This analysis is based on a draft genome for SteelHaze, however, and a complete genome would be better to accurately determine taxonomic classification.
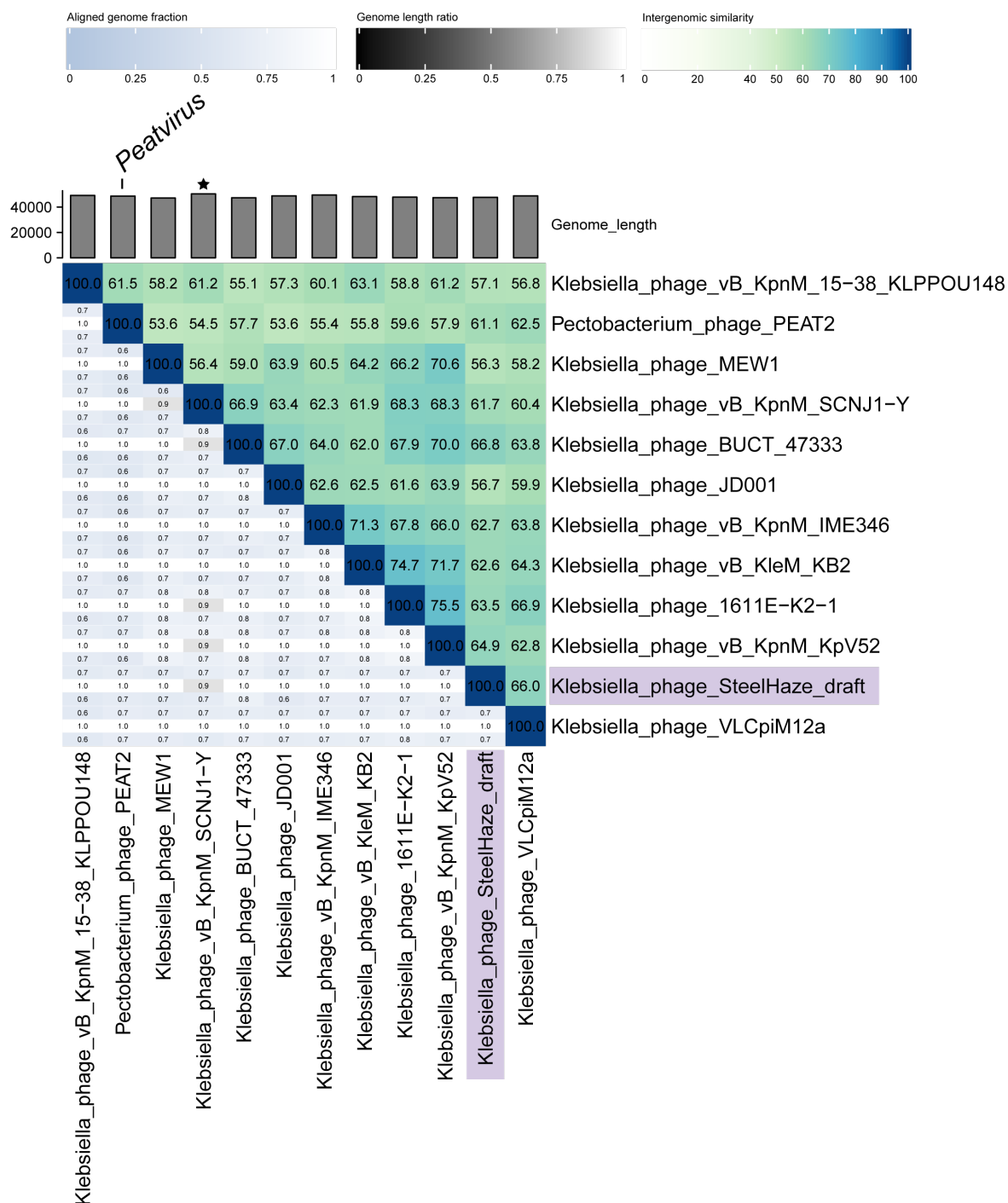
Figure 6.21. Comparative genomic analysis of SteelHaze and a selection of related unclassified myoviruses. A nucleotide-based intergenomic similarities heatmap given as percentages (right half, blue–green heatmap). The aligned genome fraction for each phage pair in the row and column is represented by the top and bottom values, respectively (left, blue-grey scale). The genome length ratio for each phage pair is represented by the middle value (grey scale). Phage SteelHaze is highlighted (purple boxes). One phage is representative of the genus *Peatvirus*. Star denotes unknown taxonomy. The other nine relatives are representative of the genus *Jedunavirus*. Heatmap generated with VIRIDIC.

6.3.5 The unclassified siphoviruses

The genome sizes of *Klebsiella* phages Iolaire and Fioreun were 48,335 and 48,342 bp, respectively, both with a GC content of 56.39%, which closely resembled the GC content of their *K. quasipneumoniae* host (~55-57%) (**Table 6.5**). The number of CDS for phages Iolaire and Fioreun was 64 and 65, respectively, and were encoded bidirectionally. Of the CDS for Iolaire and Fioreun, 33 (~51%) and 34 (~52%) were hypothetical, respectively.

Phages Iolaire and Fioreun were isolated from the same wastewater sample, from the same wastewater plant against the same *K. quasipneumoniae* strain, whereas Dereham was isolated from a different wastewater treatment plant against a *K. variicola* strain (**Appendix 10.3**). To explore their genomic relationship, a nucleotide-based intergenomic similarity analysis was performed, with Iolaire and Fioreun in the context of their closest database relatives and *Drexlerviridae* phages (see **Figure 6.14** in **section 6.3.1**). The intergenomic relationship of Dereham and its closest database relatives was explored separately (**Figure 6.22**).

Phages Iolaire and Fioreun clustered closely together and were within 98.5% nucleotide similarity across their genome length. They also clustered closely with their respective closest database relatives, *Klebsiella* phages EKq1 and P61_2 (**Table 6.5**), demonstrating ~97-98% similarity. Phages Iolaire and Fioreun were slightly less similar to *Klebsiella* phages vB_KleS-HSE3 (~92%) and vB_Kp3 (~76%), and most dissimilar to *Enterobacter* phage ATCEA85 (~49%). Moreover, Iolaire and Fioreun were considerably dissimilar to the *Drexlerviridae*, showing only 1.1-1.7% similarity. Based on these data, Iolaire and Fioreun should be assigned the same new species, within the same new genus as the relatives analysed, except *Enterobacter* phage ATCEA85. This group of phages have yet to be assigned a genus and family as they are currently assigned as unclassified siphoviruses, having been extracted from the obsolete family *Siphoviridae*.

The genome of Dereham was 47,376 bp and had a GC content of 57.03%, representing the highest GC content of any phage in the collection, and closely resembling the GC content of its host *K. variicola* (~55-57%) (**Table 6.5**). The number of CDS for Dereham was 76, despite having a smaller genome than Iolaire and Fioreun. Moreover, the CDS of Dereham were encoded unidirectionally, were more densely packed with considerably more overlap between ORFs, possibly explaining the higher number of CDS than Iolaire and Fioreun. Of the CDS, 49 were hypothetical meaning that ~64% of proteins encoded in the genome were of unknown function.

When a BLASTn search was performed for *Klebsiella* phage Dereham, only five other phages (excluding metagenome assembled genomes) were found that were ≥ 60% identity over at least 20% of the genome length, and could be included in the comparative genomic analysis (**Figure 6.22**). Phage Dereham clustered with its closest database relative

Klebsiella phage vB_Kpn_K21lambda1 (**Table 6.5**) with 83.2% similarity but was more distantly related to *Pseudomonas* phage PSA28 (69.2%) and *Salmonella* phage PMBT28 (65.3%), and highly dissimilar to *Providencia* phage Kokobel2 (23%). These data suggest that *Klebsiella* phage Dereham should be assigned a new species, within the same new genus as *Klebsiella* phages vB_Kpn_K21lambda1 and VLCpiS6a.
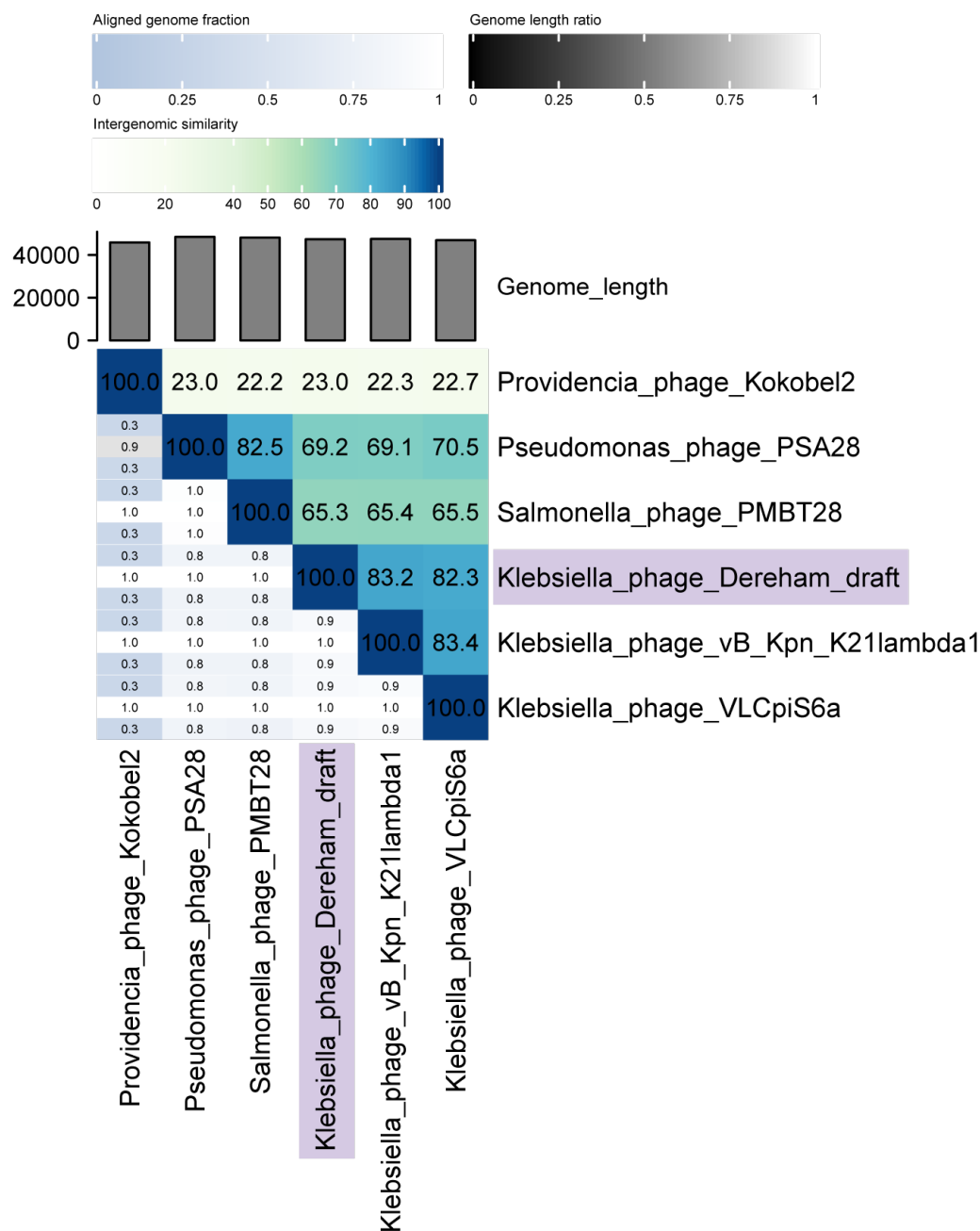
Figure 6.22. Comparative genomic analysis of Dereham and a selection of related unclassified siphoviruses. A nucleotide-based intergenomic similarities heatmap given as percentages (right half, blue–green heatmap). The aligned genome fraction for each phage pair in the row and column is represented by the top and bottom values, respectively (left, blue-grey scale). The genome length ratio for each phage pair is represented by the middle value (grey scale). Phage Dereham is highlighted (purple boxes). Heatmap generated with VIRIDIC.

6.4 Discussion

The assembly and analysis of genomic data is still challenging due to the presence of sequencing errors, repeat regions, and poor sequencing coverage (Bharti and Grimm, 2021). One of the most problematic aspects of using short-read data for phage assembly (for both *de novo* short-read-only assembly and part of a traditional hybrid assembly) was that short-read assemblers can introduce large structural errors, particularly indels in repeat regions. For the przondoviruses, such indels manifested as deletion of the right terminal repeats, possibly because the assemblers deem them as a sequencing artefact (Wick *et al.*, 2021; Wick and Holt, 2022). A second type of error that routinely occurred during non-HYPPA phage assembly was the introduction of other short indels that were particularly noticeable in coding regions. Seven of the ten przondoviruses assembled with short-read-only data contained insertions of repeat regions completely unrelated to the DTRs, but were close to, or contained within the left terminal repeat region. In fact, these erroneous repeats (~79 bp) were only identified serendipitously during the primer walking and Sanger sequencing validation of Tokugawa.

Similarly to using short-read data for polishing, a traditional hybrid assembly using both short- and long-read data in the assembly itself for Toyotomi also introduced large deletions in repeat regions, with assembly errors persisting, as has been described previously (Wick *et al.*, 2017; Wick *et al.*, 2021).

Many of the sequencing and/or assembly errors are associated with repeat regions, and/or in homopolymer regions. Specific errors are associated more often with a specific sequencing technology. For example, long-read assemblies tend to introduce more errors, particularly in homopolymer regions, but can resolve structural errors and repeat sequences that short-read sequencing cannot (Adewale, 2020; Bharti and Grimm, 2021; Wick and Holt, 2022; Wick *et al.*, 2023). Indeed, there was an argument for stricter trimming and filtering of short-read data, since long-read data were used for scaffolding phage genomes and correcting large structural errors, whereas short-read data was used for additional polishing using per-base quality thresholds.

Whilst HYPPA was successful for many phages whose sequencing data was relatively good quality, it was unable to resolve the genome of Fifoon. Three different short-read sequencing runs were performed, as inability to assemble the genome was initially hypothesised to be a sequencing rather than an assembly issue. Multiple assemblies were generated through trial and error, with short-read-only data failing to resolve the genome into a single contig. Moreover, subsampling short-read data, the application of different short-read and long-read assemblers, different polishing programs, and different combinations of polishing all failed to resolve the genome to a standard that required minimal manual curation.

One possible solution could be to generate a consensus sequence of Fifoon using Trycycler, before attempting the polishing iterations (Wick *et al.*, 2021). The use of a metagenome assembly tool could also provide an alternative method to generate a high-quality genome for Fifoon. Metagenomes are collections of multiple different genomes within a sample, but assembling phages from metagenomes has been problematic due to inability of the computational tools to capture the diversity of phages, resulting in inaccurate and poor quality metagenome-assembled phages (Mallawaarachchi *et al.*, 2023).

A possible explanation for the high sequence heterogeneity observed for Fifoon could be that it has a highly mutable DNAP. Such a DNAP would generate sequence heterogeneity that is representative of a population of Fifoon, rather than a single genome. This could explain why the genome could not be resolved for Fifoon, but could for Bolond, with which it shared high nucleotide similarity across the genome length (96.5%). However, at the time of writing, no such heterogeneity was found amongst the published literature, though quasispecies has been described in RNA viruses (Domingo and Perales, 2019; Lu *et al.*, 2020). Ironically, attempts to analyse the DNAP of the Fifoon assemblies could not be performed due to potential errors in the CDS of the DNAP.

Accurate phage genomes are essential to the use of phages as therapeutics, where they allow links between genotype and phenotype to be established, particularly with regards to providing a mechanistic understanding of infection biology. Moreover, accurate, high-quality genomes are necessary to understanding the nucleotide and proteomic structure, as well as classification of phages. Indeed, phages that are genetically similar to one another are generally avoided during combination therapy, where competition for phage receptors may result in antagonistic interactions. For example, it was established that ten przondoviruses were genetically similar to one another, and would not be useful in combination with one another due to their narrow host range, unless two distinct hosts were the target.

Comparative genomic analyses found that 25 of 26 phage genomes were likely novel species (at the time of writing), with SteelHaze likely representative of a novel genus. The intergenomic distance thresholds can be met by some phages and not others, and where thresholds are more ambiguous would require further phylogenetic analyses to confirm the correct taxonomic assignments. In particular, phages close to the genus and/or species threshold of 95% across the genome length could be incorrectly assigned (Turner *et al.*, 2021).

Little was known about the genome organisation of the unclassified *Caudoviricetes*, Iolaire, Fioreun, Dereham, and SteelHaze. Whilst there were plenty of relatives deposited in the NCBI database for some, little structural or functional information has been published, particularly regarding targeted host receptors (Maffei *et al.*, 2021). For example, no packaging mechanism could be determined for SteelHaze since it was too divergent from its closest relatives without further analyses. Packaging mechanisms can be determined for

some phages from genomic data through phylogenetic analysis of the terminase genes (*terL* and *terS*) (Merrill *et al.*, 2016). Packaging mechanisms, the type of infection cycle a particular phage employs, and genome organisation are essential to the use of phages as therapeutics, emphasising the essentiality of accurate, high-quality genomes. Currently, much of this information is missing from published literature, together with identification of primary and secondary receptors (Maffei *et al.*, 2021). Such data can provide further information on genotype-phenotype trends, which would better inform the clinical application and efficacy of phages, particularly if they are to be used in combination with other phages and/or antibiotics.

There were several limitations to the work described within this chapter. First was the type of sequencing technologies used for the HYPPA workflow. For example, the Illumina DNA Prep (a Nextera-based) kit was used for all short-read sequencing library preparations, due to their low cost and low DNA input requirements (Owen *et al.*, 2021). However, the kit does not always allow the capture of defined phage termini, since it is transposome-based and therefore needs enough DNA on both sides to be able to insert (Kot *et al.*, 2014; Owen *et al.*, 2021). Whether the short-read assemblers would have still deemed intact DTRs sequenced using a different short-read library prep kit as sequencing artefacts is yet to be fully elucidated. Regardless of the type of kit used, the issue of defined ends being assembled in the middle of the genome would persist. ONT was capable of addressing this problem since the genomes are ligated with adapters, without the use of transposomes for DNA fragmentation. For well-characterised phages whose genome organisation is highly conserved, the assembly of DTRs in the middle of the genome would not necessarily be problematic. Novel phages, for which little is known regarding their genome organisation, it would be more difficult to determine the presence or absence of defined termini, including the length and sequence of potential DTRs without the use of an assembly methodology such as HYPPA.

ONT was utilised as the long-read sequencing technology for all phages as it is more accessible and cost effective than PacBio, making it the more relevant long-read sequencing technology of the two. In an attempt to validate the HYPPA workflow using data acquired using different short-read and long-read platforms, a search of publicly-available raw data for phages within the family *Autographiviridae* was performed. Nine publicly-available phages were assembled using PacBio data, but none of the raw reads were available.

Whilst all programs utilised by the HYPPA workflow are freely available, a second limitation is the cost of both long-read and short-read sequencing and bioinformatic capabilities of processing such data that many other laboratories may not have access to.

6.5 Conclusions

This chapter described the process of trial and error in generating complete, accurate, and high-quality phage genomes. Producing such genomes was more difficult than initially anticipated, with persistent sequencing or assembly errors found amongst traditional methods of phage genome assembly. To overcome these difficulties, the HYPPA workflow was developed and validated using a case study of ten highly conserved genomes of phages belonging to the genus *Przondovirus* that require minimal manual curation, even for more difficult to resolve genomes. Moreover, HYPPA was successful in resolving the genomes of another eight phages in the collection. At the time of writing however, ten phage genomes were still designated as draft due to long-read sequencing failures.

Classification of all genomes (whether draft or complete) found a diverse range of phages belonging to seven genera across four known families. Three unclassified *Caudoviricetes* represented novel species, and one represented a novel genus. However, future work will include further analyses to generate high-quality genomes and confirm the taxonomic assignment of draft genomes and submission of a taxonomy proposal for further classification of novel genomes. Further genomic and functional studies may also be necessary to determine packaging mechanisms of novel phages and identify primary and/or secondary host receptors.

CHAPTER 7: TRANSPOSON MUTAGENESIS IN *KLEBSIELLA MICHIGANENSIS*

7.1 Introduction

Transposon-directed insertion site sequencing (TraDIS) was first developed to determine the genes involved in bile tolerance that contributed to *Salmonella enterica* serovar Typhi carriage (Langridge *et al.*, 2009). TraDIS allows the high-throughput assay of multiple genes simultaneously through the creation of single-gene mutants within a pooled library (Langridge *et al.*, 2009; Holden *et al.*, 2021). Traditional TraDIS and other transposon-directed mutagenesis methods rely on transposon insertions in genes, meaning that whilst it can be used to identify essential genes, it cannot be used to assay them (Yasir *et al.*, 2020; Holden *et al.*, 2021). The transposon mutant library can be used to determine genes involved in the bacterial response to specific conditions, by sequencing of the inserts to identify the frequency and location of the transposon and comparing them to the non-stressed control (Turner *et al.*, 2020; Holden *et al.*, 2021). Recently, TraDIS-*Xpress* has been developed, which makes use of transposon-located outwards-transcribing promoters, which allow deeper investigation into how gene expression, as well as inactivation, affects survival under a stress (Yasir *et al.*, 2020; Holden *et al.*, 2021). This also allows essential genes to be assayed, which cannot be assayed using conventional TraDIS (Yasir *et al.*, 2020; Holden *et al.*, 2021). Since its conception, TraDIS has been used to understand genes involved in survival, stress responses, antimicrobial tolerance, virulence, and even phage infection in a range of pathogenic bacteria (Jana *et al.*, 2017; Yasir *et al.*, 2020; Bruchmann *et al.*, 2021; Holden *et al.*, 2021; Acton *et al.*, 2024).

Phages are obligate parasites capable of hijacking the host replicative machinery to produce phage progeny. Infection requires adsorption through interaction with host primary and secondary receptors in a reversible and irreversible manner, respectively (Guttman *et al.*, 2005; Bertozzi Silva *et al.*, 2016; Maffei *et al.*, 2021). Despite their discovery over a century ago, there is still relatively little known about phages and the molecular mechanisms by which they infect their hosts outside of model phages, such as T4 and T7 (Guttman *et al.*, 2005). Particularly with encapsulated bacteria such as *Klebsiella* spp., the exact nature and distinction between initial interaction with the bacterial cell, especially the CPS, and irreversible attachment has yet to be fully elucidated (González-García *et al.*, 2015; Latka *et al.*, 2017; Maffei *et al.*, 2021). That is, it is unclear whether the capsule is a barrier for which phages need to overcome, or whether attachment to the capsule constitutes the reversible attachment step. Indeed, many phages targeting *Klebsiella* spp. encode depolymerases that degrade CPS and/or biofilms, which are necessary for access to bacterial surface receptors and infection (Cai *et al.*, 2019). For some phages, specific receptors are well-characterised. For example, LPS components are almost always identified as the secondary receptor in Gram-negative-targeting podoviruses (Molineux, 2006; Maffei *et al.*, 2021), with CPS components possible primary receptors.

Transposon mutagenesis has been conducted in *Klebsiella*, with most being in *K. pneumoniae*, but to our knowledge, a TraDIS library has not been constructed in *K. michiganensis*. To investigate phage host receptors and any other genes involved in the bacterial response to phage infection, a TraDIS mutant library was constructed in a clinical strain of *K. michiganensis*. Identifying these genes would facilitate the design of a phage cocktail targeting *Klebsiella* spp. by combining phages targeting distinct receptors.

This chapter will examine the construction of the TraDIS mutant library and identification of genes involved in bacterial response to phage infection, with results and discussion combined for each section.

7.1.1 Aims

1. Construct a TraDIS mutant library in *K. michiganensis* M7 21 2 #35 WT
2. Use the TraDIS mutant library to determine genes important during phage infection
3. Design a phage cocktail using these data, together with the phage infection kinetics data

7.2 Construction and optimisation of *Klebsiella michiganensis* TraDIS mutant library

To determine bacterial genes involved in phage infection on a large-scale, a Tn5 transposon (Tn5+Kan^R) insertion mutant library was constructed in *K. michiganensis* M7 21 2 #35 WT (hereafter *K. michiganensis* M7 WT). This strain was selected due to the number of phages in the collection that were able to productively infect it.

7.2.1 Transposon mutagenesis in *Klebsiella*

Several factors inherent to *Klebsiella* (and some other genera of Gram-negative bacteria) had to be considered before constructing a mutant library. The first one was the capsule that typifies the *Klebsiella* genus can significantly reduce transformation efficiency *in vitro* (Haudiquet *et al.*, 2021; Wannigama *et al.*, 2023). The second was the type I restriction-modification (RM) system, which degrades foreign DNA upon entering the bacterial cell and would be able to degrade the transposon DNA before it integrated into the genome, and can also reduce transformation efficiency.

To address both factors, a small-scale TraDIS mutant library was constructed with different concentrations of EDTA (to remove/reduce capsule) (Wannigama *et al.*, 2023) and presence/absence of a type I restriction inhibitor (to prevent the RM system from acting on the transposon) at a concentration of 5 µg per 50 µL of cells.

Experimental treatment conditions included EDTA added to both the overnight culture and the culture grown to exponential phase on the day of the experiment (double-dose), with or without the restriction inhibitor and were designated:

- Double-dose of EDTA at 0.7 mM, 0.7+0.7 (+/- restriction inhibitor)
- Double-dose of EDTA at 70 mM, 70+70 (+/- restriction inhibitor)

Double-dosed conditions: cells were grown overnight with concentrations of EDTA at 0 mM, 0.7 mM, and 70 mM. On the day of the experiment, two groups of cells (50 µL) were prepared and grown to exponential phase with concentrations of EDTA at 0 mM, 0.7 mM, and 70 mM. restriction inhibitor was added to the second group of cells only.

Additionally, experimental treatment conditions included EDTA added to only the culture grown to exponential phase on the day of the experiment (single dose), with or without restriction inhibitor and were designated:

- Single dose of EDTA at 0.7 mM, 0.7 (+/- restriction inhibitor)
- Single dose of EDTA at 70 mM, 70 (+/- restriction inhibitor)

Single dose conditions: cells were grown overnight without EDTA. On the day of the experimental, two groups of cells (50 µL) were prepared with concentrations of EDTA at 0 mM, 0.7 mM, and 70 mM, with restriction inhibitor added to the second group of cells only, creating six experimental conditions. A further four samples received a double dose of EDTA added to the overnight culture.

The transformation efficiency was higher overall in the cells that did not receive the restriction inhibitor compared to the cells that did (**Figure 7.1**). Moreover, the double-dose of EDTA was inversely proportional to transformation efficiency (**Figure 7.1**). Whilst transformation efficiencies were not statistically significant within each group, there was a statistically significant difference (p = 0.0429) between 0.7 mM EDTA without restriction inhibition compared to 0.7 mM EDTA with restriction inhibition (**Figure 7.1**). Based on these data, the experimental conditions selected for the main TraDIS experiment were no EDTA in overnight culture, 0.7 mM EDTA added before transformation, and no addition of restriction inhibitor.

Construction of the TraDIS mutant library was subsequently performed in two batches. During the washing steps of the first batch, pellets were resuspended and pooled (see **Chapter 2**) prior to electroporation. To investigate whether the density of cells affected the transformation efficiency during the small-scale trial, the cells from the first batch were pooled, whereas the second batch were diluted 1:1 in ice-cold 10% glycerol prior to transformation. For both groups of cells, the recovery time was also increased from 1:30 to 2:30 hours.
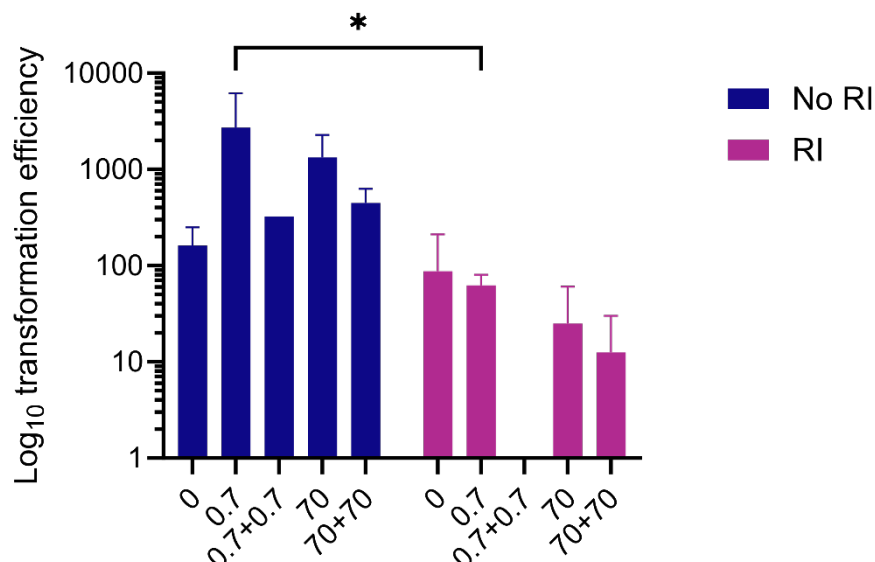
Figure 7.1. Transformation efficiency of _K. michiganensis_. Media was supplemented with 50 µg/mL kanamycin and different concentrations of EDTA (0 mM, 0.7 mM, 70 mM) alongside doubled-dose EDTA (0.7 mM or 70 mM added to overnight culture in addition) with and without restriction inhibition (RI). Data represent mean ± SD of two technical replicates. Differences between EDTA concentrations with and without restriction inhibition were analysed by two-way ANOVA corrected for multiple comparisons (Tukey method). _p_ ≤ 0.05 (∗).

The colonies from the pooled and un-pooled populations were enumerated. Approximately 9,000 transformants were recovered per plate from the dense cell population, whereas approximately 35,000 transformants were recovered per plate from the diluted cell population. Transformation efficiencies were calculated as 2.33×10^8 and 6.00×10^7, respectively, with a fold change of approximately 3.9 between the dense and diluted cell populations.

Based on the number of calculated transformants across the two batches, plus those from the trial run, approximately 400,000 _K. michiganensis_ transposon mutants had been generated.

7.2.2 Addressing the challenges of transposon mutagenesis in _Klebsiella_

Sequencing of the _K. michiganensis_ TraDIS mutant library was performed prior to phage infection experiments to assess the quality of the library and the distribution of the insertions across the genome. Following read mapping to the genome assembly, very few of the reads mapped back to the original _K. michiganensis_ M7 WT reference genome (data not shown). The reads, however, did map to the plasmid from which the Tn5+Kan[R] had been amplified. This suggested that the majority of mutants did not contain a chromosomally-integrated transposon but were in fact contained within the plasmid. This suggested that either the

transposon was not excised effectively from the plasmid (by linearisation of the Tn5+Kan[R], or by the transposase) and the mutants were the result of plasmid replication, or that the transposon was excised from the plasmid correctly, but reintegrated within very small quantities of carryover plasmid, and then replicated *in situ*. Thus, while 400,000 mutants had been generated, very few contained insertions within chromosomally-encoded genes in *K. michiganensis* (described below). Due to time constraints, no further library construction was attempted.

To determine the position and number of unique insertion sites (UIS) generated by transposon mutagenesis, primers were used to sequence out from the Tn5 transposon element of the *K. michiganensis* TraDIS mutant library containing all the pooled mutants (as described by Yasir *et al.* (2020)). The sequence downstream of the transposon was used to identify the locus at which the transposon inserted, and these loci were mapped onto the *K. michiganensis* M7 WT reference genome. The number of UIS describes the number of unique loci where the transposon has inserted through the whole genome. The number of UIS were normalised to the gene length because the number of reads that map to a particular position in a gene is dependent upon the gene length (Langridge *et al.*, 2009). Comparing the insertion frequencies (both the UIS and the abundance of reads mapped to each locus) between control and phage-treated conditions identified genes that affect susceptibility to phage treatment. Before comparisons were made however, the number of UIS had to be determined for the *K. michiganensis* TraDIS mutant library.

The first sequencing run of the mutant library showed approximately 6,300 unique insertion sites across the bacterial genome, equating to one insertion every 1,000 base pairs (**Figure 7.2**). With a genome size of 6,259,885 bp and approximately 5,800 genes, there would be one insertion per gene. However, a second independent replicate subset of the library was sequenced, which showed approximately 4,800 unique insertion sites, equating to 0.86 insertions per gene. Moreover, the reference *K. michiganensis* M7 WT was a draft genome consisting of 31 contigs, and contig overlap would result in an overestimation of transposon insertions that spanned multiple contigs. As transposon mutagenesis would knock out essential genes, for which it is not possible to assay, these data would suggest that not every gene had been hit, but that on average, almost every non-essential gene would have at least one mutant.

Some transposons show bias towards AT-rich regions and some for GC-rich regions. The TraDIS mutant library did not appear to show bias for transposon insertion within such regions, but did show potential transposon insertion hotspots irrespective of AT- or GC-richness, particularly at approximately 1,000,000 bp (contig four), for example (**Figure 7.2**). Because the genome is in separate contigs and not fully assembled, the GC content is not evenly distributed, and no replication origin and terminus could be determined. There is a skew between the number of GC nucleotides due to the mechanism of DNA replication,

which is semi-conservative (Grigoriev, 1998; Hubert, 2022). DNA is synthesised from a leading strand and a lagging strand where the nucleotides are not evenly distributed between the two strands (Grigoriev, 1998; Hubert, 2022). Guanine (G) and thymine (T) are more abundant in the leading strand, whereas cytosine (C) and adenine (A) are more abundant in the lagging strand, creating a GC skew, which can then indicate the replication origin and terminus, for which many bacteria such as *Klebsiella* spp. have only one (Harding *et al.*, 1982; Grigoriev, 1998; Hubert, 2022). Moreover, the GC skew can indicate recombination events due to phage infection, for example (Grigoriev, 1998).
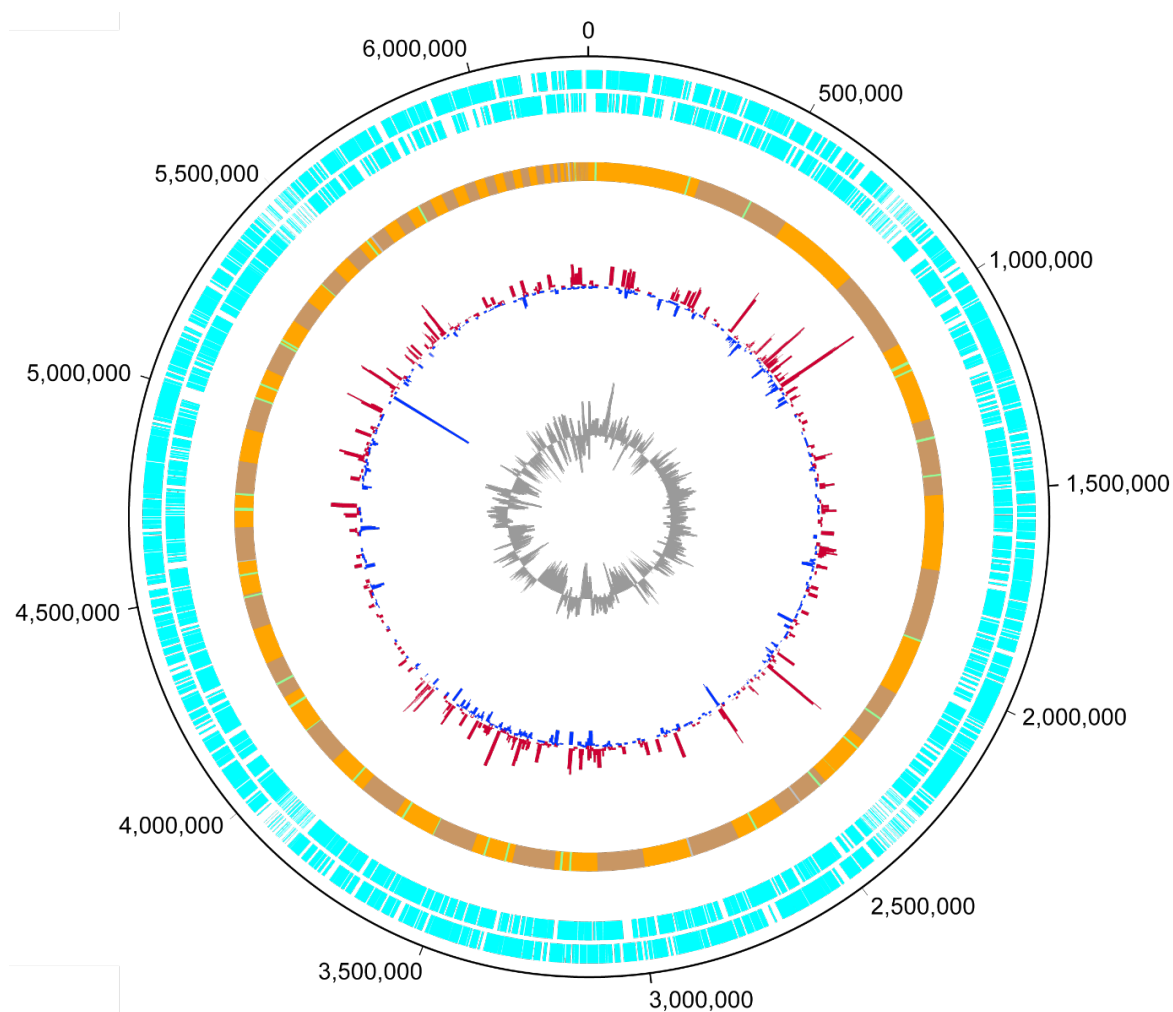
Figure 7.2. Insertion frequency of transposon across the *K. michiganensis* genome. Insertion frequency on the forward (red) and reverse (blue) strands. Inner track shows GC content (grey). Orange track shows the genome contigs with green lines showing tRNA genes. Genes on the forward and reverse strands are shown in cyan on the outer and inner tracks, respectively. Figure generated with DNAPlotter.

7.2.3 Designing the phage infection TraDIS experimental conditions

To determine the genes involved during phage infection, a preliminary screening of phages against the TraDIS mutant library was performed by plaque assay and in liquid culture.

Preliminary screening was performed with phages based on the time-kill assays: phage Oda was identified as a fast-replicating, whereas Fifoon and Assynt were identified as slow-replicating (see **Chapter 4**). Spotting of both phages on a lawn of *K. michiganensis* TraDIS mutant library and the *K. michiganensis* M7 WT as a control, and incubation overnight revealed a number of mutants in the area of phage clearance, indicating phage-resistant mutants. The number of mutants were enumerated for the TraDIS mutant library for both phages and compared to the WT control (**Figure 7.3**).
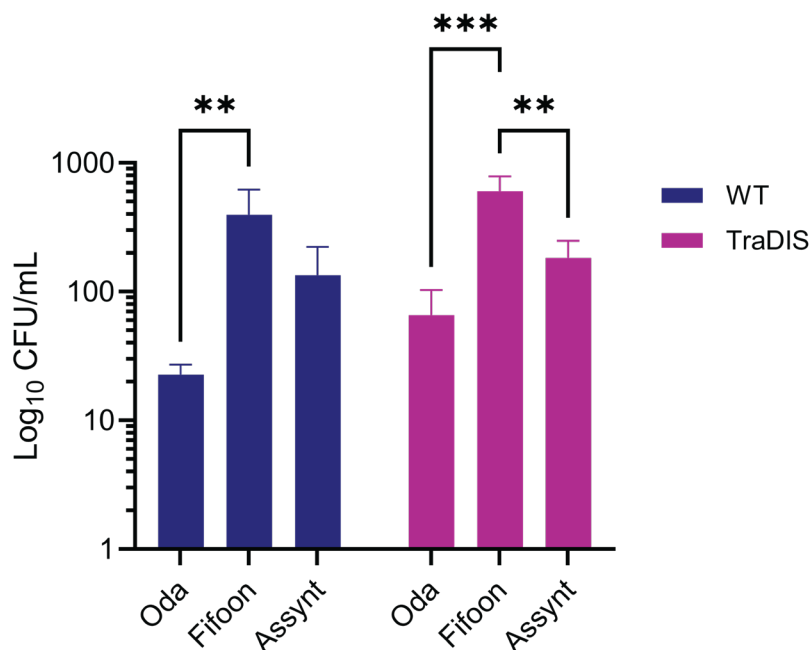
Figure 7.3. Screening of TraDIS mutant library with three different phages. Comparison of the number of mutants generated following overnight incubation of *K. michiganensis* M7 WT (purple) and *K. michiganensis* TraDIS mutant library (purple) with phages Oda (*Przondovirus*), Fifoon (*Slopekvirus*), and Assynt (*Jiaodavirus*). Data represent mean ± SD of three technical replicates. Differences in CFU/mL between the different phages against the two bacterial conditions were analysed by two-way ANOVA corrected for multiple comparisons (Tukey method). $p \leq 0.01$ (∗∗); $p \leq 0.001$ (∗∗∗).

More mutants were identified when each phage was tested against the TraDIS mutant library compared to the WT, but were not statistically significant ($p > 0.05$) (**Figure 7.3**). These data suggested that there were marginally more phage-resistant mutants in the TraDIS mutant library when compared to the mutants in the WT strain by infecting the host with high-titre phages. Interestingly, more mutants were observed for Fifoon than Oda against the WT (**Figure 7.3**). Similarly, more mutants were observed in the TraDIS mutant library for Fifoon than both Oda and Assynt (**Figure 7.3**). Possible explanations for the differences between the phages could represent differences in their targeted host receptors and their infection dynamics. For example, slower-replicating phages may allow more time for the emergence of phage-resistant mutants. The lack of statistically significant difference between the WT and TraDIS mutants could be contributed by the rapidity with which such *de novo* mutations arise following incubation overnight with high phage titres and the low

density of the TraDIS library. However, these data are semi-quantitative at best, and the preliminary screen gave a very rudimentary overview of whether the TraDIS mutant library did indeed have enough mutants to perform phage challenge experiments.

Next, the TraDIS mutant library was challenged in liquid culture with Oda and SteelHaze, representative of fast- and medium-replicating phages, respectively, that were isolated against and capable of infecting the parent strain *K. michiganensis* M7 WT (run 1). Two time points were selected at 1:30 and 3:00 hours for both phages and at a single MOI of 10. Given that Oda likely completes its lifecycle within approximately 20 minutes, it was necessary to establish whether the MOIs and time points were deemed too long. Following incubation, phage Oda had almost completely collapsed the bacterial stock at both time points whereas SteelHaze did not. DNA was extracted for both the control and the library following challenge by both phages at both time points and the transposon insertion sites were sequenced.

Following sequencing however, the number of reads that mapped back to the *K. michiganensis* M7 WT genome was very low (between 0.3 and 3%) after infection by both phages, whereas the number of reads for the control at both time points was ≈1.1% (**Appendix 10.16**). The number of UIS identified at 1:30 hours post-inoculation (hpi) with phage Oda, which was just 10, suggested that at the MOI and the time point were too high to enrich for a significant number of mutants for which any data could be extracted (**Figure 7.4A**, **Appendix 10.16**). The number of UIS for phage Oda at 3:00 hpi however, was high (> 2,000) (**Figure 7.4A**, **Appendix 10.16)**. The number of UIS for both the control (> 4,000) and following challenge with SteelHaze (> 3,000) at the same MOI as Oda was high (**Figure 7.4A**, **Appendix 10.16**). Based on the data for Oda at 1:30 hpi however, a decision was made to lower the MOI to 1 for further experiments for all phages and time points to maintain consistency.

The TraDIS mutant library was challenged with five phages in duplicate at an MOI of 1 (run 2). The five phages tested were Bolond, Fifoon, Oda, SteelHaze, and Eggie, and the same two time points as tested previously. Bolond and Fifoon were selected to see whether there would be differences observed in the gene hits in the TraDIS mutant library when challenged with two highly similar phages (96.5% intergenomic nucleotide similarity) from the same genus. Similarly, Eggie was selected as a genetically diverse phage (12% intergenomic nucleotide similarity) from the same family. Oda was selected as the type species of the family *Autographiviridae* within the phage collection, as well as SteelHaze, which is the first representative of a novel genus.
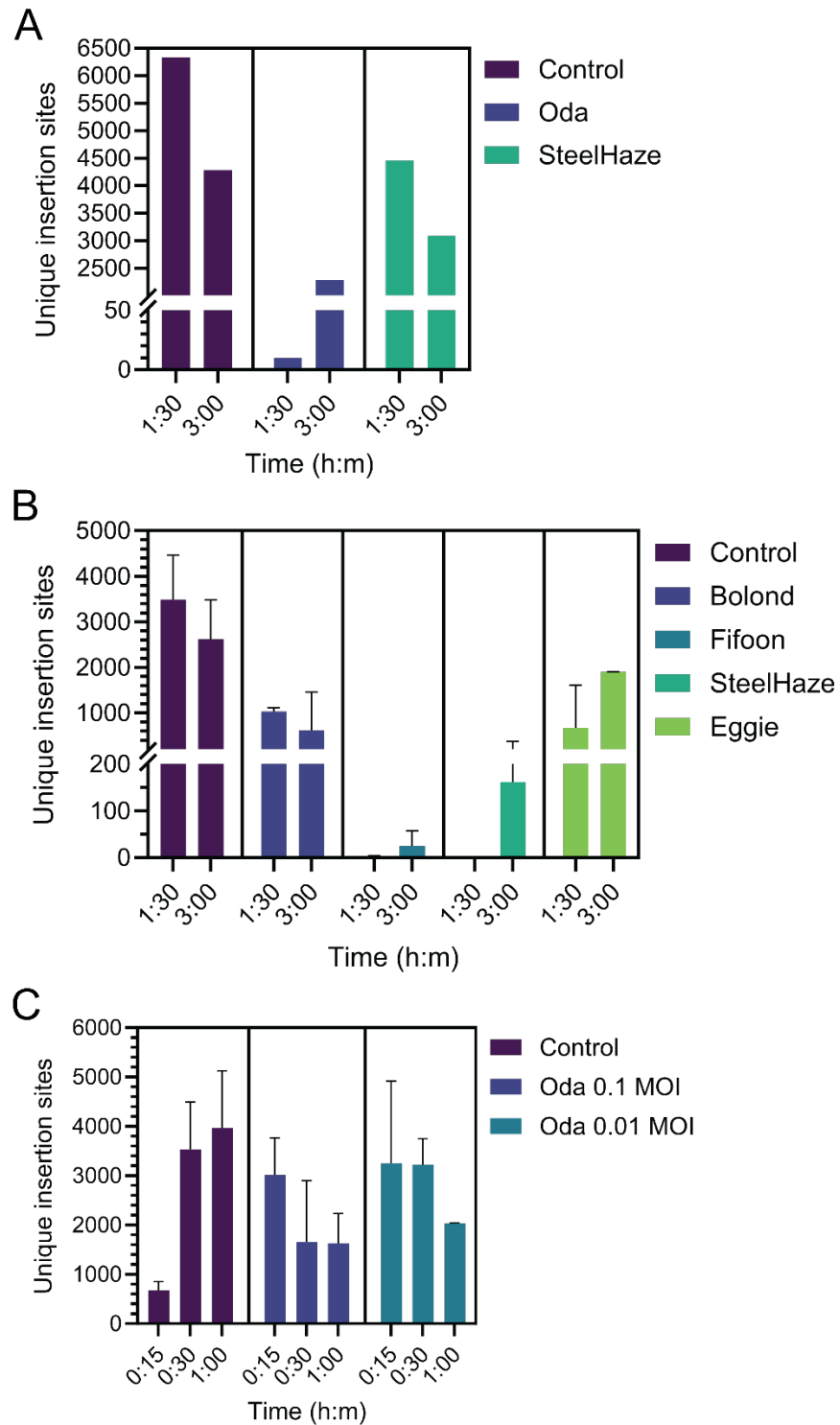
Figure 7.4. Number of unique insertion sites normalised to total sequence length in *K. michiganensis* TraDIS mutant library following phage challenge. (A) The first TraDIS experiment (run 1) for no phage control and phages Oda and SteelHaze at two different time points. Phage MOI = 10. Data represent a single replicate from a single run. **(B)** The second TraDIS experiment (run 2) for no phage control and all phages except Oda at two different time points. Phage MOI = 1. No data was determined for SteelHaze at 1:30 hpi. Data represent two technical replicates from a single run. **(C)** The third TraDIS experiment (run 3) for no phage control and phage Oda at two different time points and two different MOIs. Data represent two technical replicates from a single run.

In run 2, Oda completely collapsed the bacterial stock again, making it impossible to extract sufficient DNA for sequencing (≈3 ng/µL). The TraDIS mutant library was sequenced following challenge with the remaining phages (Bolond, Fifoon. SteelHaze, and Eggie), with differences amongst the number of UIS identified (**Figure 7.4B**, **Appendix 10.17**). Despite a high number of reads for most replicates at both time points, challenge with Bolond and Fifoon resulted in fewer than 1,250 and 48 total UIS, respectively (**Figure 7.4B**, **Appendix 10.17**). Moreover, only one replicate at 1:30 hpi could be sequenced following challenge with SteelHaze, and all three replicates across the two time points tested resulted in fewer than 310 UIS (**Figure 7.4B**, **Appendix 10.17**). Challenge with Eggie resulted in up to 1,900 UIS except for one replicate, which had only two (**Figure 7.4B**, **Appendix 10.17**). The control group had between 2,000 and 4,200 UIS for both time points tested (**Figure 7.4B**, **Appendix 10.17**).

To address the issue of phage Oda collapsing the bacterial stock, the TraDIS mutant library was challenged at two lower MOIs (0.1 and 0.01) and three different time points (0:15, 0:30, and 1:00 hpi) in run 3. Considerably more UIS were generated at all time points, with more in general observed at MOI 0.1 than at 0.01 (**Figure 7.4C**, **Appendix 10.17**). Considerably fewer UIS were observed for the control group at 0:15 hpi than for the 0:30 or 1:00 hpi (**Figure 7.4C**, **Appendix 10.17**).

7.3 Identification of genes involved in response to phage infection

TraDIS was used to determine genes involved during phage infection following challenge with five phages when compared to a non-phage treated control. The TraDIS mutant library was fragmented and size selected (see **Chapter 2**) and the fragments containing the Tn5 were enriched and sequenced. The sequencing reads were mapped back to the *K. michiganensis* M7 WT genome, where the location and frequency of the transposon mutants were determined.

Differences in the number of insertions in a particular gene can inform its relative fitness under the specific conditions. Increased frequency of insertions in a particular gene following phage infection relative to the control (i.e. control has fewer insertions) would suggest mutants have increased fitness under these conditions (Cain *et al.*, 2020; Acton *et al.*, 2024). That is, the genes are considered detrimental to bacterial survival during phage infection, and deletion/disruption of such genes are beneficial to the bacteria. Such genes would be phage receptors or other genes essential for phage infection, and are designated a susceptibility gene. These genes would show a positive $\log_2$ fold-change (logFC). Putative susceptibility genes in all plots are indicated in orange.

Reduced frequency of insertions in a particular gene following phage infection relative to the control (i.e. control has more insertions) would suggest mutants have reduced fitness under these conditions (Cain *et al.*, 2020; Acton *et al.*, 2024). That is, the genes are considered essential or beneficial to bacterial survival during phage infection, and deletion/disruption of such genes are detrimental to the bacteria. Such genes could be involved in anti-phage defence, for example, and are designated a resistance gene. These genes would show a negative logFC. Putative resistance genes in all plots are indicated in purple.

Gene hits have been plotted as logFC (y-axis) against -$\log_{10}$ of the q-value (x-axis), which represents statistical significance corrected for false discovery rate. A major caveat to these set of experiments is that because there were so few insertional mutants in the TraDIS mutant library, there is little statistical power in the q-value. Thus, all plots are representative of potential susceptibility and/or resistance genes and cannot be considered statistically significant. For this reason, points within the arbitrary logFC parameters (≤ -3.0 for resistance gene hits and ≥ 3.0 for susceptibility gene hits) for which there is a q-value of zero, have still been plotted. Data for phage Fifoon at 3:00 hpi and SteelHaze at 1:30 hpi could not be plotted as the pipeline failed to produce summary statistics, likely because of the low number of reads. This is despite fewer reads generated for Fifoon at 1:30 hpi (**Appendix 10.17**).

7.3.1 Identification of genes involved in phage resistance

Initial observation found that more genes potentially involved in phage resistance were identified than susceptibility genes following infection with phages Bolond, Fifoon, SteelHaze and Eggie (**Figure 7.5A-D**).

The top hits for potential resistance genes following phage infection by Bolond (**Figure 7.5A**), Fifoon (**Figure 7.5B**), SteelHaze (**Figure 7.5C**), and Eggie (**Figure 7.5D**) during run 1 were proteins often associated with the outer membrane. For example, *yesR* and *ccpA* (Bolond, **Figure 7.5A**) are involved in the catabolism of L-rhamnose in *Bacillus subtilis*, which is part of a larger rhamnose catabolism operon, with orthologous operons in *E. coli* and *S.* Typhimurium (Rodionov, 2007; Hirooka *et al.*, 2016). Rhamnose is a plant- and soil-associated sugar (Hirooka *et al.*, 2016), but has also been found in the CPS and EPS of some *Klebsiella* spp. (Patro and Rathinavelan, 2019). Other resistance gene hits included *wcaJ* (Fifoon and SteelHaze, **Figure 7.5B** and **Figure 7.5C***, respectively) and *wzyE* (Eggie, **Figure 7.5D**), which are involved in CPS and ECA biosynthesis, respectively (Patro *et al.*, 2020; Rai and Mitchell, 2020; Lam *et al.*, 2022). Other potential resistance genes included those encoding enzymes involved in nutrient metabolism. For example, *hcaB* was also identified in four of the five phages tested, and encodes an enzyme involved in phenylpropionic acid catabolism in *E. coli* as an alternative carbon source in soil (Díaz et al., 1998).
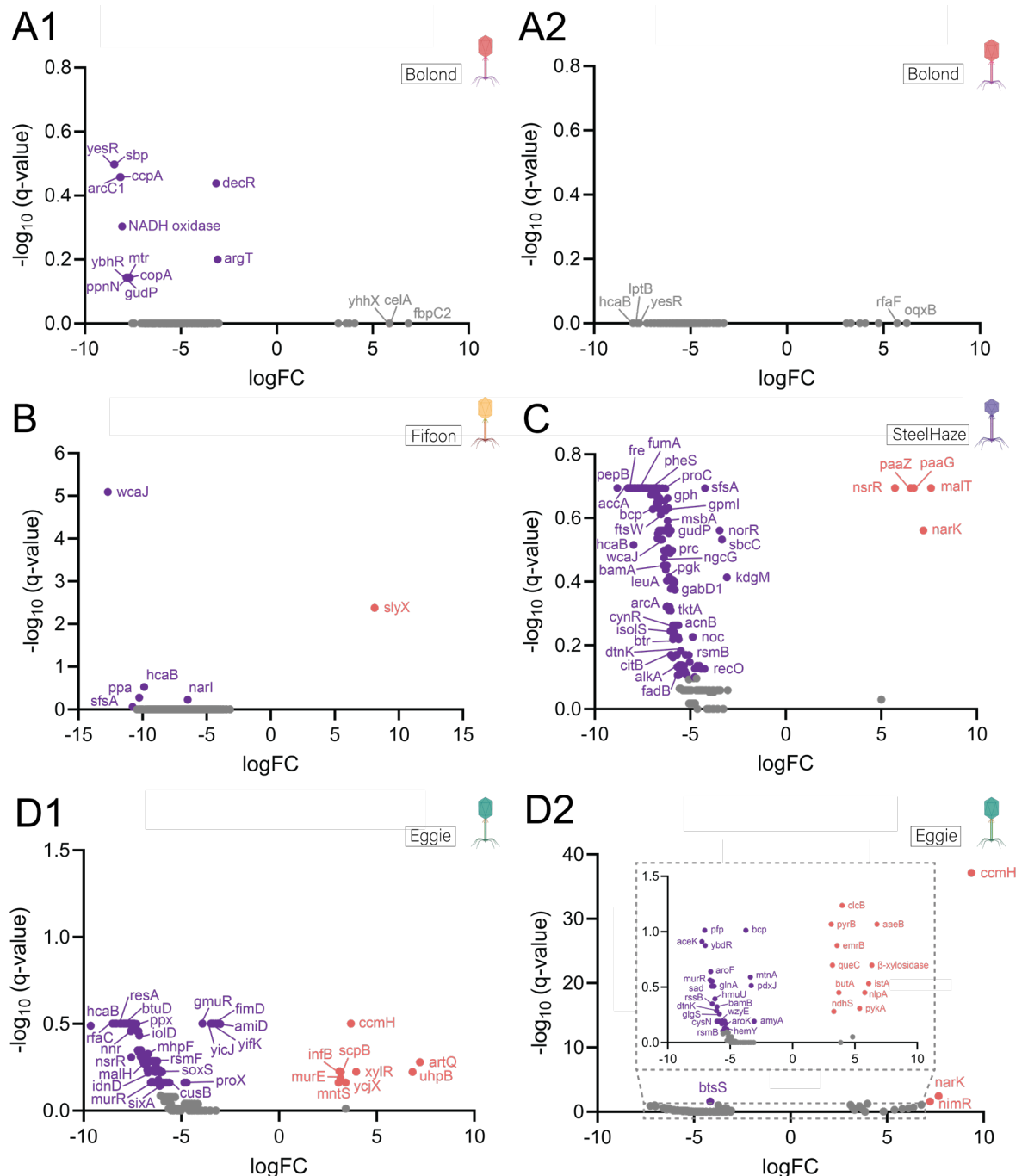
Figure 7.5. Differences in transposon insertion within *K. michiganensis* genes following infection by four different phages. Log$_2$ fold change (logFC) of all gene hits with statistical significance (-log$_{10}$ of the q-value). Gene hits following infection by phage **(A1)** Bolond at 1:30 hpi and **(A2)** 3:00 hpi; **(B)** Fifoon at 1:30 hpi only; **(C)** SteelHaze at 3:00 hpi only; and **(D1)** Eggie at 1:30 hpi and **(D2)** 3:00 hpi with inner panel of gene hits up to -log$_{10}$ (q-value) 1.5. Genes that showed fewer insertions (≤ -3.0 logFC) relative to the control are given in purple, and genes that showed greater insertions (≥ 3.0 logFC) are given in orange. Hits within the logFC parameters but with a -log$_{10}$ (q-value) of 0 are given in grey. Hits between -2 and 2 logFC were not included. All phages were tested at MOI = 1.

Other genes identified as potential resistance genes included those involved in LPS biosynthesis (*rfaC*, Eggie) (Kwon *et al.*, 2016); cell wall biosynthesis (*murR*, Eggie) (Jaeger and Mayer, 2008; Weber *et al.*, 2020); vitamin B12 transport (*btuD*, Eggie) (Pieńko and Trylska, 2020; Maffei *et al.*, 2021; Acton *et al.*, 2024); and the citric acid cycle and oxidative

stress (*fumA* and *aceK*, SteelHaze; *soxS*, Eggie) (Park and Gunsalus, 1995; Rodionov, 2007; Anes *et al.*, 2021); pyruvate metabolism (*btsS*, Eggie); and sugar fermentation (*sfsA*, Fifoon and SteelHaze; *malH*, Eggie) (Takeda *et al.*, 2001) (**Figure 7.6B-D**). Contrastingly, Acton *et al.* (2024) identified LPS biosynthesis and vitamin B12 genes as putative susceptibility genes when a *S.* Typhimurium TraDIS mutant library was challenged with four phages from three distinct genera.

Infection by phage Oda was performed separately at two different MOIs and at three different time points (**Figure 7.6A-B**). Initial observation found that more genes potentially involved in phage resistance were identified than susceptibility genes following infection with phage Oda at 1:00 hpi for both 0.1 (**Figure 7.5A3**) and 0.01 MOIs (**Figure 7.6B**), but not other time points.

Top hits for potential resistance genes following challenge with Oda included those observed for other phages, such as *wzyE*, *hcaB*, and *malH*. Potential resistance genes specifically identified following challenge at 0.1 MOI with phage Oda included *nasD* and *yedS* at 0:30 hpi (**Figure 7.6A2**) and 1:00 hpi (**Figure 7.6A3**), respectively. In addition to *narI* and *narK* (discussed in **section 7.3.3**), *nasD* is also involved in nitrate metabolism, as it encodes part of a membrane-bound nitrate transporter (Wu and Stewart, 1998). YedS, a protein encoded by *yedS*, is not usually translated in *E. coli* expressing the OmpF and OmpC porins (OmpK35 and OmpK36 homologues in *Klebsiella*, respectively) (James *et al.*, 2009; Tsai *et al.*, 2011; Warner *et al.*, 2013). Downregulation, deletion, and or truncation of OmpK35 and OmpK36 in *K. pneumoniae* contributes to carbapenem resistance (Pagès *et al.*, 2008; Tsai *et al.*, 2011; Wong *et al.*, 2019; Wong *et al.*, 2022). However, Warner *et al.* (2013) found that downregulation of OmpF and OmpC in *E. coli* resulted in the upregulation of AcrAB-TolC efflux via expression of the *marRAB* operon, Moreover, the *marAB* operon regulates the expression of *yedS*, which has been implicated in carbapenem resistance in the absence of functional OmpK35 and OmpK36 porins. These data suggest the potential for cross-resistance to carbapenems following infection by phage Oda.

Other potential resistance genes identified following infection with phage Oda included those associated with biofilm and fimbriae regulation (*glgS*, *pdeG*) (Povolotsky and Hengge, 2015; Joudeh *et al.*, 2021); ABC transporters, for example LPS (*msbA*) (Kwon *et al.*, 2016; Alexander *et al.*, 2018) and amino acid (*gltI*) transport (Niu *et al.*, 2023); CPS biosynthesis (*galF*) (Lam *et al.*, 2022); and proteins involved in osmoregulation (*mdoH*) (Debarbieux *et al.*, 1997). Most were observed following phage Oda challenge at 1:00 hpi for both 0.1 MOI (**Figure 7.6A3**) and 0.01 MOI (**Figure 7.6B3**).

Additionally, resistance genes associated with transcriptional regulation of multifactorial proteins (*cynR*, *nagR*) (Rodionov, 2007); iron scavenging and oxidative stress (*sufB*) (Anes *et al.*, 2021); and inactivation of antimicrobial peptide (*ptrB*) were identified at 1:00 hpi for 0.01 MOI only (**Figure 7.6B3**). Similar to the response to infection with Eggie, genes within

cell wall biosynthesis (*murD*) (Weber *et al.*, 2020) and citric acid cycle operons (*fumD*) were also implicated as potential resistance genes (**Figure 7.6B3**). Moreover, inactivation mutants of *fumD* in *E. coli* can confer resistance to aminoglycoside antibiotics such as streptomycin (Vlasblom *et al.*, 2014).

Based on these data, potential mechanisms of resistance could include modification or loss of LPS and capsule, and genes associated with nutrient uptake, which have been reported previously (Maffei *et al.*, 2021; Acton *et al.*, 2024).

7.3.2 Identification of potential phage receptors

Fewer genes potentially involved in phage susceptibility were identified than resistance genes following infection with phages Bolond, Fifoon, SteelHaze and Eggie (**Figure 7.5A-D**). However, following challenge with phage Bolond at 1:30 hpi no susceptibility gene hits were statistically significant, and at 3:00 hpi no hits for either resistance or susceptibility gene hits were statistically significant. Whilst not statistically significant, hits identified as potential susceptibility genes and thus potential phage receptors for phage Bolond did include those associated with antibiotic resistance (*oqxB*, *yhhX*) (Li *et al.*, 2019); and LPS biosynthesis (*rfaF*) (Kwon *et al.*, 2016) (**Figure 7.5A**). *yhhX* mutants have been implicated in resistance to chloramphenicol in *E. coli* (Duo *et al.*, 2008), suggesting that mutants could thus confer cross-resistance following phage infection.

The *slyX* gene was the only potential susceptible gene hit following infection by Fifoon at a single time point (**Figure 7.5B**). No known function has been assigned this gene in *Klebsiella*, but is possibly involved in phage phiX174 infection in *E. coli* (Roof *et al.*, 1994).

paaZ and *paaG* genes were identified as potential susceptibility genes following challenge with phage SteelHaze (**Figure 7.5C**), both of which are part of a gene cluster that has been implicated in iron scavenging in *Pseudomonas* (Wang *et al.*, 2023). A gene involved in nitrate metabolism, *narK* was also implicated as a susceptibility gene following challenge with SteelHaze (**Figure 7.5C**) and Eggie at only 3:00 hpi (**Figure 7.5D1**).

Conversely to *murR*, which was identified as a resistance gene following infection with Eggie, another mediator of cell wall biosynthesis, *murE* (Weber *et al.*, 2020) was identified as a susceptibility gene for the same phage at the same time point (**Figure 7.5D1**). Additionally, *ccmH* was identified as a potential susceptibility gene following infection by Eggie at both time points (**Figure 7.5D**). This gene is also involved in the electron transport chain and iron acquisition (Fabianek *et al.*, 1999; Kwon *et al.*, 2016). Other potential susceptibility gene hits following infection by Eggie included those encoding proteins involved in translation (*infB*) and pyrimidine biosynthesis (*pyrB*) (Weber *et al.*, 2020). Other gene hits included *clcB*, *artQ*, and efflux pumps (*aaeB*) (Jindal *et al.*, 2019) (**Figure 7.5D**). Interestingly, *infB* is a housekeeping gene and would typically be considered an essential gene (Diancourt *et al.*, 2005; Herzog *et al.*, 2014; Kwon *et al.*, 2016).
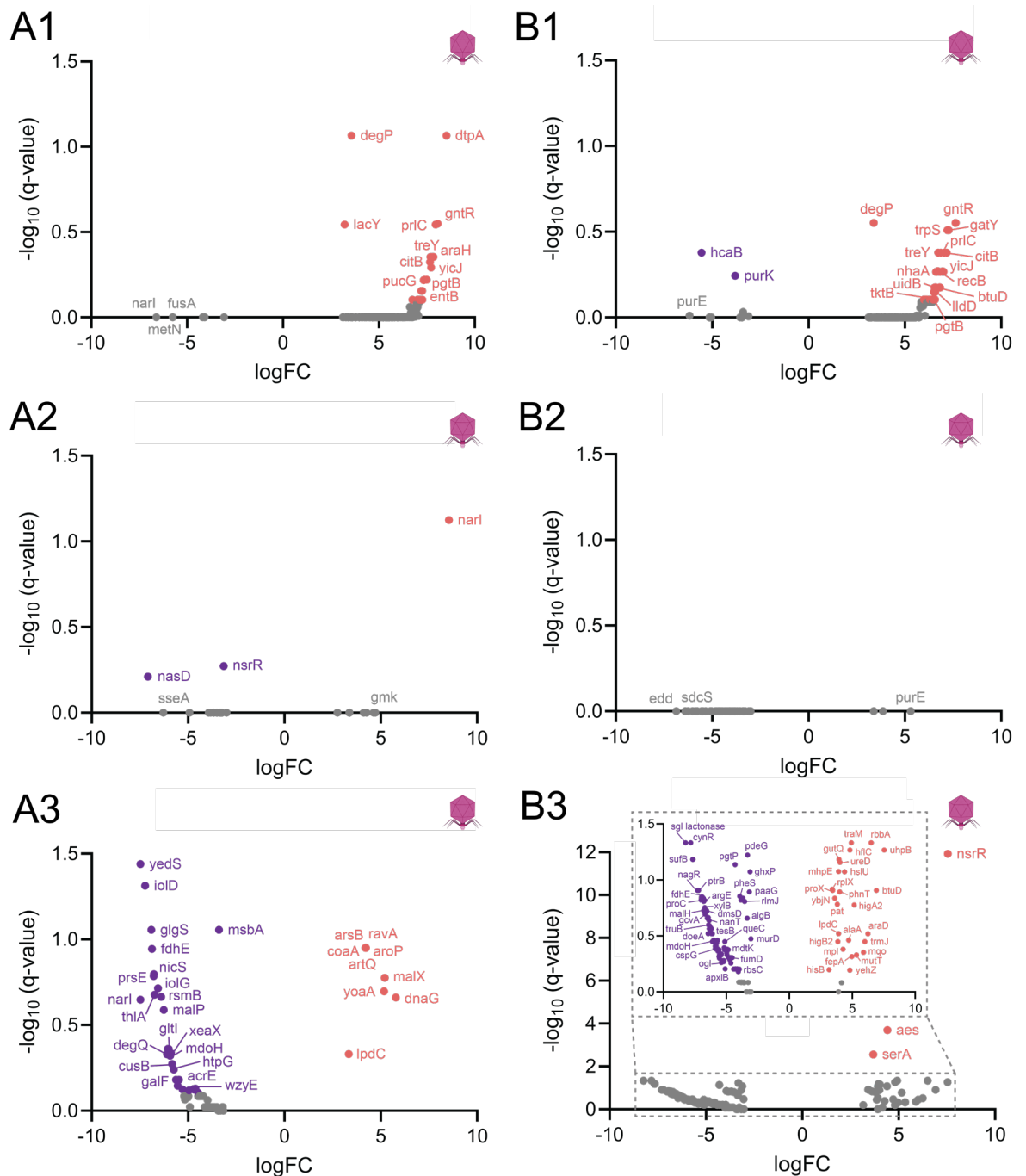
Figure 7.6. Differences in transposon insertion within _K. michiganensis_ genes following infection by phage Oda. Log₂ fold change (logFC) of all gene hits with statistical significance (-log₁₀ of the q-value). Gene hits following infection by phage Oda at MOI = 0.01 at **(A1)** 0:15 hpi, **(A2)** 0:30 hpi, and **(A3)** 1:00 hpi. Gene hits at an MOI = 0.1 at **(B1)** 0:15 hpi, **(B2)** 0:30 hpi, and **(B1)** 1:00 hpi with inner panel of gene hits up to -log₁₀ (q-value) 1.5. Genes that showed fewer insertions (≤ -3.0 logFC) relative to the control are given in purple, and genes that showed greater insertions (≥ 3.0 logFC) are given in orange. Hits within the logFC parameters but with a -log₁₀ (q-value) of 0 are given in grey. Hits between -2 and 2 logFC were not included.

treY, _degP_, _citB_ and _gntR_ were identified as susceptibility genes following challenge with Oda at 0:15 hpi for both MOIs (**Figure 7.6A1**, **Figure 7.6B1**). Additionally, _btuD_ was identified as a susceptibility gene at 0.1 MOI at 0:15 hpi (**Figure 7.6A1**) and 1:00 hpi (**Figure 7.6A3**). _treY_ may be important in managing environmental stress (Carpinelli _et al._, 2006),

whereas *degP* is important in stress resistance in *E. coli* (Zhang *et al.*, 2019). *citB* is a protein important in the metabolism of citrate in *K. pneumoniae*, but may also be implicated in transcriptional regulation of the SOS response following exposure to β-lactam antibiotics (Scheu *et al.*, 2012). *gntR* is a transcriptional regulator involved in many essential cellular processes such as metabolism, virulence, and antimicrobial resistance (Rodionov, 2007; Guzmán *et al.*, 2011; Paczosa *et al.*, 2020; Liu *et al.*, 2021).

Based on these data, susceptibility genes (i.e. phage receptors) could include factors associated with cell wall and LPS biosynthesis, AMR such as efflux pumps, factors associated with virulence, such as iron scavenging, and factors associated with stress responses.

7.3.3 Genes with differential insertional frequency

Interestingly, genes associated with nitrate metabolism (*narI*, *narK*) were differentially hit depending on the phage, the MOI, and/or the time point.

narI was identified as a potential resistance gene following challenge with Fifoon (**Figure 7.5B**). However, *narI* was identified as a potential resistance and susceptibility genes under different MOIs and times points following challenge with phage Oda (**Figure 7.6**). Specifically, *narI* was identified as a potential resistance gene at 0:15 hpi (**Figure 7.6A1**) and 1:00 hpi (**Figure 7.6A3**) at 0.1 MOI, but fell below the plotting thresholds at both time points (logFC 0.427 and -1.631, respectively). At the 0:30 hpi however, *narI* was identified as a susceptibility gene at the 0.1 MOI only (**Figure 7.6A2**), whilst at 0.01 MOI, *narI* was not observed following infection by phage Oda at any time point. *narK* was identified as a susceptibility gene following infection by Eggie at only one of the two time points however (**Figure 7.5D**).

Nitrate reductase genes (*narG*, *narH*, *narI*, and *narJ*) are involved in anaerobic respiration: *narI* encodes a subunit of the nitrate reductase protein, whereas *narK* encodes a nitrate/nitrite transporter protein, both of which are found in the bacterial inner membrane and are important in the electron transport chain (Sodergren *et al.*, 1988; Wu and Stewart, 1998; Moreno-Vivián *et al.*, 1999; Kwon *et al.*, 2016). Moreover, these genes have been identified as virulence genes in *K. pneumoniae* strains (Kwon *et al.*, 2016).

Another gene exhibiting differential expression was *nsrR*. *nsrR* insertional mutants were identified as a potential susceptibility gene following challenge with SteelHaze (**Figure 7.5C**), but as a potential resistance gene following challenge with Eggie at 1:30 hpi (**Figure 7.5D1**) only. Similarly, *nsrR* insertional mutants were differentially implicated as both resistance and susceptibility genes, but at different MOIs and time points following challenge with phage Oda. Specifically, it was identified as a potential resistance gene at 0.1 MOI after 0:30 hpi (**Figure 7.6A2**), but a susceptibility gene at 0.01 MOI after 1:00 hpi (**Figure 7.6B3**). The *nsrR* gene is a multifactorial transcriptional repressor involved in

regulation of stress responses in *E. coli* (Filenko *et al.*, 2007; Partridge *et al.*, 2009). Differential insertional frequency may suggest differences in gene expression during the phage stress response between phages SteelHaze, Eggie, and Oda.

Insertional mutants involved in purine biosynthesis (*purK*, *purE* and *proC*) were also identified as potential resistance genes in Oda, whereas *purE* was identified as a potential susceptibility gene at 0.1 MOI 0:30 hpi only.

Following challenge with five phages from four distinct genera, TraDIS identified many potential resistance and susceptibility genes, including those with differential insertional frequency depending on the phage, MOI, and/or time point. The three main genes identified that exhibited differential insertional frequency were *narI*, *narK*, and *nsrR*. Possible differences between time points, such as *narK* insertional mutants at 3:00 hpi with Eggie but not 1:30 hpi could be that the mutation is not particularly beneficial at the earlier time point. Indeed, at an MOI of 1, phage Eggie had not reduced bacterial load below the LOD until 2:00 hpi (see **Figure 4.5** in **section 4.4.3**). This suggested that at 3:00 hpi, *narK* insertional mutants were enriched compared to the control, and were essential to bacterial survival in the presence of phage for this time point only, whereas at 1:30 hpi, the essentiality of the gene was not as high.

The differential insertional frequency of *narI* following infection by Fifoon and Oda suggested that mutational inactivation decreased bacterial fitness following infection Fifoon, but increased bacterial fitness following infection by Oda only at 0.1 MOI 0:30 hpi. However, at 0.1 MOI 0:15 and 1:00 hpi, *narI* insertional mutants were enriched compared to controls. These data suggested that *narI* played differential roles for both phages, and a differential role when challenged with Oda at different time points. Similarly, *btuD* was identified as a resistance gene following infection by Eggie at 1:30 hpi, but as a susceptibility gene following infection by Oda 1:00 hpi at 0.01 MOI only.

Differential insertional frequency has recently been shown to result in collateral sensitivity of phages targeting *S.* Typhimurium (Acton *et al.*, 2024). For example, mutations in the LPS biosynthetic pathway conferred bacterial resistance to one phage increased sensitivity to another phage (Acton *et al.*, 2024). The mechanism of action was hypothesised to be loss of the O-antigen receptor which was masking a vitamin B12 transporter, leading to bacterial resistance to the first phage, but increased sensitivity to the second phage (Acton *et al.*, 2024).

Interestingly, some phages exhibited differences in the number of insertional mutants at particular time points. For example, following challenge with phage Oda, more susceptibility genes were identified at 0:15 hpi for both MOIs, whereas more resistance than susceptibility genes were identified at 1:00 hpi for 0.01 MOI. A possible explanation could be that phage Oda completes its life cycle within 20 minutes, whereby mutants that confer increased

bacterial survival are more likely to be enriched at 0:15 hpi, whereas the opposite is true as the incubation period continues.

Although three phages from the same family (Bolond, Fifoon, and Eggie) were used to challenge the TraDIS mutant library, incomplete data and the low number of mutants in the library made it difficult to draw any conclusions regarding potential host receptors. Moreover, whether any of the phages could be used in combination based on their predicted receptors is yet to be fully elucidated. The differences in insertional mutant frequency suggested that specific genes are important in bacterial fitness following infection, which is dependent not only to the type of phage infecting, but also dependent on the phage titre at specific points in time. The lack of a highly dense TraDIS mutant library was a major limitation however, particularly as none of the data presented here had any statistical power. Moreover, none of the results could be validated through single-gene knockouts, and thus any conclusions are speculative at best.

7.3.4 Future work

To validate any results obtained from the phage challenge of the TraDIS mutant library, genome editing of *K. michiganensis* M7 WT by homologous recombination (recombineering) was planned to create single-gene knockouts. The preliminary experiments to prepare the host strain for recombineering were performed, but no knockouts were generated due to time constraints.

In preparation for recombineering two plasmids were required and extracted from two parent strains: pSIM18 from *Salmonella enterica* subsp. *enterica* serovar Typhimurium SL144, and pKD4 from *E. coli* BW25141.

K. michiganensis M7 WT cells were transformed with pSIM18 using the protocol outlined in **section 2.2.4** without issue and plated at 30°C with hygromycin supplementation before storage at -80°C.

Future work should include further optimisation to improve transformation efficiency, with the aim of generating a library of at least 400,000 mutants, and validation of any results through challenging single-gene knockouts with phage. Current lab work by others includes the investigation of the plasmid carry-over with steps taken to get rid of the plasmid before transformation, optimisation of the transposome incubation times and optimisation of the transformation conditions to improve efficiency.

7.4 Conclusions

A *K. michiganensis* TraDIS mutant library was constructed, and the library was challenged with five different phages to identify genes required for phage infection. A transposon mutant library of approximately 6,000 mutants was created, equating to approximately one insertion

every 1,000 bp. Due to time constraints however, optimisation to improve the transformation efficiency in constructing the *K. michiganensis* TraDIS mutant library was not possible.

Optimisation of the TraDIS experiments for phage challenge included adjusting time points and MOIs, which was based on the data from the kinetics experiments as well as trial TraDIS experiments.

With the approximate 6,000 mutants in the library, many of the gene hits associated with susceptibility and resistance were identified as membrane-associated factors, such as transporters and their regulators, and nutrient metabolism. However, the findings had no statistical power due to the size of the mutant library.

The preliminary data from this chapter, together with the phage infection dynamics assays, suggest that combinations of three phages, Oda, SteelHaze, and Eggie would make the best candidates for a phage cocktail, but more data is needed to confirm they do not target the same receptors on the bacterial cell surface.

CHAPTER 8: OVERALL DISCUSSION

8.1 Overall perspectives

A diverse and well-characterised *Klebsiella* isolate collection was built as the basis for the bacteriophage work in this project. The isolates in the physical collection were recovered from a range of different sources, including clinical, food, environmental, and type strains. The collection spans most *Klebsiella* species, with *K. pneumoniae* the most commonly isolated species from clinical and food samples.

This project identified 26 genetically distinct phages that were isolated against six different *Klebsiella* hosts spanning five species. Phages were isolated from various river water and wastewater samples at different times and locations across England. These novel phages spanned five known genera and three new genera within four families: two phages in the genus *Slopekvirus* and nine in the genus *Jiaodavirus* (both in the family *Straboviridae*); ten phages in the genus *Przondovirus* (in the family *Autographiviridae*); one phage in the genus *Webervirus* (in the family *Drexlerviridae*); three phages in two new genera of unclassified siphoviruses and a new genus of unclassified myovirus in the class *Caudoviricetes*.

Phages were characterised phenotypically to assess their host range, killing capabilities in liquid culture, and to explore potential phage-antibiotic synergy. These data put into context the suitability of their use for future therapeutic applications, in combination with other phages from the collection as well as antibiotics. The most efficacious phage combinations were those from three distinct lineages; Oda (*Autographiviridae*), SteelHaze (unclassified *Caudoviricetes*), and Eggie (*Straboviridae*), which demonstrated differences in their killing dynamics against the same *K. michiganensis* host. Combination dosages were tailored based on the killing dynamics, with phage Oda at a ten-fold dilution to the other two, due to its rapid infection cycle. The combination of genetically distinct phages has previously been shown to increase the spectrum of activity (i.e. increased host range) with a concomitant reduction in resistance emergence (Merabishvili *et al.*, 2018). Moreover, the three-phage cocktail combined two narrow spectrum phages and one broad spectrum phage. Generally, narrow spectrum phages would have limited clinical value for a defined phage product, unless combined with broad spectrum phages and/or antibiotics, which have broad spectrum activity relative to phages (Abedon *et al.*, 2021). Additionally, T-even phages such as Eggie have previously been shown to utilise multiple receptors, especially in the absence of their "preferred" receptor (usually porins), and their adaptability thus makes them good candidates for phage cocktails (Guttman *et al.*, 2005; Merabishvili *et al.*, 2018; Kortright *et al.*, 2020; Maffei *et al.*, 2021).

The data presented here showed that phage-antibiotic combinations were highly dependent on the phage used in the combination. Indeed, phages can demonstrate vast differences in efficacy when combined with antibiotics with the same mechanism of action and is highly dose-dependent (Akturk *et al.*, 2019). For example, Akturk *et al.* (2019) found that higher

concentrations well beyond the MIC of the antibiotic gentamicin did not lead to a greater anti-biofilm effect when combined with phage, likely because gentamicin targets the protein synthesis machinery upon which phages rely to replicate (Krause *et al.*, 2016). Moreover, gentamicin led to a greater anti-biofilm effect when compared to other protein synthesis inhibitors, possibly because it is rapidly bactericidal when compared to other antibiotics in the same class, such as kanamycin (Krause *et al.*, 2016; Akturk *et al.*, 2019). Moreover, the authors demonstrated that phages in combination with a single antibiotic applied together exhibited a moderate efficacy when compared to sequential application (Akturk *et al.*, 2019). That is, phage appeared to clear biofilm, which allowed access of the antibiotic to the cell surface, increasing its diffusion across the OM, demonstrating greater killing (Akturk *et al.*, 2019). Prudent selection of phage-antibiotic combinations has also been highlighted by Gu Liu *et al.* (2020), who showed different classes of antibiotics exhibited different behaviour when combined with the same phage. For example, the bactericidal antibiotics cefotaxime and ciprofloxacin target cell wall synthesis and DNA replication, respectively, and exhibited different behaviours when combined with SteelHaze (synergistic and additive, respectively). Moreover, different phages when combined with the same antibiotic also exhibited different behaviours. For example, chloramphenicol was synergistic when combined with phage SteelHaze, but antagonistic when combined with phage Eggie. Phage-antibiotic combinations have also been shown to require reduced antibiotic dosing (Gu Liu *et al.*, 2020). This could have important implications for use of toxic antibiotics, such as colistin, which is used as a last resort antibiotic due to its nephrotoxicity (Sabnis *et al.*, 2021).

A recent study reported the successful use of a pre-adapted phage together with antibiotics in the treatment of a PDR *K. pneumoniae* infection of a bombing survivor (Eskenazi *et al.*, 2022). Both clinical and genomic data reported were extensive, and the phage used in the study was classified as belonging to the genus *Slopekvirus* in the family *Straboviridae*. Such use of a well-characterised phage in this study meant that assumptions could be made regarding the host receptors, and any future phage combinations could be made accordingly. Moreover, genomic analyses revealed a missense mutation in the hinge connector of the distal tail fibre in the pre-adapted phage, when compared to the wild-type, enabling researchers to track the evolution of the phage (Eskenazi *et al.*, 2022). This study demonstrated how clinical and genomic data together can significantly improve clinical phage therapy outcomes.

A major regulatory hurdle to the use of phages as therapeutics is the inevitable development of phage resistance of the host bacteria. During this study, we found that three-phage combination therapy was more successful at killing and suppressing regrowth than single-phage preparations, suggesting reduced selection of resistant mutants, but did not test this directly. The application of phages and the phenomenon of phage resistance can be mitigated two-fold. Firstly, resistance is energetically expensive, for both phage resistance and antibiotic resistance, where both result in a concomitant fitness cost in growth and/or

virulence (Vogwill and MacLean, 2015; Gao *et al.*, 2022). The application of phages has previously been shown to result in a fitness cost that re-sensitised the bacteria to the antibiotics to which there were previously resistant (Gao *et al.*, 2022). Moreover, fitness costs in response to phage predation can also result in reduced virulence: for example, loss of LPS resulted in reduced virulence following phage cocktail application in a mouse model (Gao *et al.*, 2022). The most important and well-characterised virulence factor amongst *Klebsiella* spp. is the capsule, and capsule loss has been demonstrated to occur as a result of phage predation. Moreover, acapsular variants are attenuated *in vivo* (discussed below). Secondly, the application of phage cocktails, with or without antibiotics has also been shown to be highly efficacious, minimising the resistance emergence to multiple factors simultaneously (Torres-Barceló *et al.*, 2014; Abedon *et al.*, 2021). Where some antibiotic combinations with a similar mechanism of action (i.e. trimethoprim and sulfamethoxazole) can result in synergy (Bollenbach, 2015; Minato *et al.*, 2018; Gu Liu *et al.*, 2020), phages with distinct mechanism of action from each other and their antibiotics, tend to be synergistic (Gu Liu *et al.*, 2020; Abedon *et al.*, 2021).

An additional concern would be the emergence of cross-resistance. Kortright *et al.* (2021) found that resistance of *E. coli* to phage T6 resulted in cross-resistance to another phage and the antibiotic albicidin due to mutations in the Tsx porin, which all three utilised to exert their effects. The LPS was a common receptor identified in another study amongst a group of phages, and LPS modification or loss resulted in cross-resistance to all phages (Adler *et al.*, 2021).

Together, the data presented in this project demonstrate careful selection of phage-phage and phage-antibiotic combinations, with extensive phenotypic characterisation that goes beyond simple phage-antibiotic synergy testing. Adequate characterisation of the co-evolutionary trade-off between phages and their hosts is necessary to understand and mitigate the regulatory concerns. Indeed, functional genomic data arising from TraDIS experiments has complemented this work by indicating likely bacterial phage receptors and resistance determinants, which would better inform combinatorial decisions.

The clinical application of phage therapy to treat recalcitrant infections often lack detailed genomic and functional data, reporting only the basic information, and in some cases, no genomic data is attributed to the phage used in the clinical study. For example, Corbellino *et al.* (2020) demonstrated the successful treatment of a recalcitrant UTI caused by *K. pneumoniae* with a phage preparation sourced from Georgia. However, no genomic data of the phage was reported.

Many phage research articles regarding genomics and/or clinical studies do not go beyond reporting basic genomic characteristics of phages, however. Moreover, functional studies linking genotype to phenotype are considerably lacking (Maffei *et al.*, 2021). Additionally, many phage genomes deposited within databases are incomplete, often containing

sequencing and/or assembly errors, as was encountered in this study with some of the reference przondoviruses. A major aim of this work was to generate complete, accurate, and high-quality phage genomes. Persistent sequencing and assembly errors were found amongst traditional methods of phage genome assembly, which were overcame through the development of the HYPPA workflow. Indeed, Eskenazi *et al.* (2022) used a traditional short-read-only assembly without any further curation or polishing steps.

The rationale for long-read-only assembly during the HYPPA workflow was to create a consensus sequence without any large structural errors, particularly in repeat regions, and correcting any errors with long- and short-read polishing (Wick and Holt, 2022; Wick *et al.*, 2023). The highly conserved genome organisation of the ten przondoviruses made testing and validation of the HYPPA workflow less challenging, especially when using well-characterised and well-curated phages such as Enterobacteria phage K30 as an exemplar.

The HYPPA workflow was also capable of resolving the genomes of a further eight genetically distinct phages in the collection with minimal manual curation. However, the HYPPA workflow was unable to resolve the genome of one phage, and was hypothesised that Fifoon exists as a population of closely related genomes, rather than a clonal population that is observed for all other phage isolates in the collection, similar to the quasispecies concept known for RNA viruses with high mutation rates (Domingo and Perales, 2019; Lu *et al.*, 2020). The cause for quasispecies could be due to hypermutation, possible due to the presence of an error-prone DNAP (Domingo and Perales, 2019; Lu *et al.*, 2020). Hypermutability has been described for multiple bacterial taxa in response to phage stress (Morgan *et al.*, 2010; Chevallereau *et al.*, 2020) and in long term evolution assays (Lenski *et al.*, 2015), but at the time of writing, no such hypermutator has been described amongst tailed phages.

Producing complete, accurate, and high-quality phage genomes are integral to their use as therapeutics. Complete functional and genomic data in particular, are important in biobanking of phages. For example, phages will inevitably mutate with subsequent rounds of propagation, and with every treatment application, thus phages should be collected, sequenced, and assembled regularly to analyse the adaptations that are taking place *in situ*. A literature search of phage biobanking initiatives found that there was currently no consensus of how frequently biobanked phages are or should be sequenced and assembled (Lin *et al.*, 2021; Nagel *et al.*, 2022; Bosco *et al.*, 2023). The work presented here significantly built on the understanding of phage genomics, and contributed to the generation of complete, accurate, and high-quality phage genomes. Phage assembly using the HYPPA workflow will greatly improve the number of high-quality phage genomes deposited in the databases, thereby providing a greater understanding of genotype-phenotype link and phage-host interactions, life cycle and packaging strategies, and their evolution over time. Moreover, accurate phage assembly will prove invaluable to biobanking

measures which are currently being implemented worldwide (Lin *et al.*, 2021; Nagel *et al.*, 2022). Producing such high-quality genomes however, was much more difficult than previously anticipated. Indeed, the sequencing and bioinformatics capabilities may not be accessible or feasible for some researchers, and would make meeting the requirements for phage therapy particularly challenging.

The *Klebsiella* capsule is often the first host-derived component that phages come into contact with. A large study exploring the complex interplay between 46 phages and their bacterial host found that the *Klebsiella* capsule determines the specificity with which phages can infect (Beamud *et al.*, 2023). Indeed, most of the phages tested in that study had a narrow host range. No link could be established between the phages characterised in this project and the ST or capsular type of their bacterial hosts. However, our collection was very diverse, spanning across the entire genus, and it is likely that finer grained effects like ST and capsular type preferences of the phages were not picked up. Moreover, the phages characterised in this project were a mix of broad and narrow host range.

This specificity has several caveats. First, the capsule being the first bacterial component phages interact with is subject to continuous adaptation. For example, two studies found that capsule loss is driven by phage predation, which leads to a concomitant increase in HGT by conjugation (Haudiquet *et al.*, 2021; Haudiquet *et al.*, 2024). Moreover, conjugative mating can restore capsule, and even promote serotype switching (Haudiquet *et al.*, 2021; Haudiquet *et al.*, 2024). These data have implications for the use of phages in the treatment or prevention of *Klebsiella* infections. In addition, acapsular variants are attenuated *in vivo* (Lawlor *et al.*, 2005; Fung *et al.*, 2011), but this capsule loss is likely temporary. However, Lourenço *et al.* (2023) have demonstrated that the use of acapsular variants of *Klebsiella* can select for capsule-independent phages, with a broader host range, overcoming this problem. For example, the use of phages in combinations can address these problems. Phenotypic and functional genomics studies have found collateral sensitivity can arise through combination therapy (Acton *et al.*, 2024). Using TraDIS, Acton *et al.* (2024) found that one particular mutant could resist infection by one phage through modification/loss of its O-antigen, previously identified as the phage receptor. This mutant exhibited collateral sensitivity to another phage however, as the loss of O-antigen facilitated access to its receptor, the vitamin B12 transporter BtuB (Acton *et al.*, 2024). Moreover, a study in *E. coli* phages have found that degradation of the capsule by phage-encoded depolymerases facilitated access of other phages in the cocktail to host receptors, in which the capsule was previously masking (Schmerer *et al.*, 2014). Capsule and biofilm degradation by phage-encoded depolymerases have also been shown to potentiate the activity of antibiotics and human immune system (Azeredo and Sutherland, 2008; Wang *et al.*, 2019). These data demonstrate the utility of phage combinations when rationally designed.

To address the lack of functional studies, transposon mutagenesis was performed together with TraDIS by massively-parallel assay of all genes involved during phage infection. TraDIS has previously been used to explore stress responses in a range of pathogenic bacteria and response to phage infection (Langridge *et al.*, 2009; Jana *et al.*, 2017; Yasir *et al.*, 2020; Bruchmann *et al.*, 2021; Holden *et al.*, 2021; Acton *et al.*, 2024). Whilst the number of TraDIS mutants generated during this project were low, the results presented here do demonstrate differences between genetically distinct phages, and interestingly, differences in genes involved at different infection time points for the same phage which was not systematically tested in the study by Acton *et al.* (2024). TraDIS has previously revealed novel proteins involved in phage susceptibility and resistance for two model *E. coli* phages, T4 and T7 (Cowley *et al.*, 2018), highlighting it as a powerful tool for functional genomics, even for phages that have been studied for decades. Moreover, functional genomic studies such as TraDIS can complement phenotypic studies to determine the most efficacious phages for use in therapy, particularly in selecting those that target distinct receptors, and make predictions on the bacterial resistance that may emerge, and even subsequent phage adaptations to evade resistance. A functional genomics study by Adler *et al.* (2021) also found that mutations in regulatory genes had significant global consequences particularly in nutrient uptake of the bacteria. Indeed, phage cross-resistance was multifactorial and often indirectly related to the phage receptor, more likely attributed to the nutritional status (Adler *et al.*, 2021). These data provide insight into why so many nutrient uptake genes were hit in the TraDIS experiments performed here. With this importance of the capsule noted above, gene hits for capsule components were found during the TraDIS experiments performed here for specific phages. However, capsule genes were not found for all phages, possibly due to the small size of the TraDIS library.

Lastly, the biggest bottleneck in producing phages as therapeutics, besides producing complete, accurate, and high-quality genomes, is producing phages to GMP (Jones *et al.*, 2023; House of Commons Committee, 2024; Suleman *et al.*, 2024). Considerable cost and infrastructure are necessary to generate large-scale production of phage preparations with standardised protocols (Jones *et al.*, 2023; Suleman *et al.*, 2024). Here, a pilot-scale phage amplification and purification workflow was implemented. Whilst amplification to large volumes and high titres was successful, downstream concentration and purification was not. These efforts indicate that phage production optimisations likely need to be made for each phage individually and this will add to the product development cost significantly.

8.2 Future work

The application of phages targeting *Klebsiella* spp. present a promising alternative or adjuvant to current antibiotic therapy. However, *Klebsiella* is a common commensal of human mucosal membranes, including the nasopharynx and gut. Indeed, human carriage of *Klebsiella* spp. is the largest risk factor for developing invasive infection. As such, future

work should explore whether *Klebsiella* eradication from the gut has any off-target effects on the human microbiome.

Future work should focus on further phage amplification and purification optimisation to generate a workflow that is scalable, standardised, and can generate high-quality, high-titre phage preparations. Particularly important, is generating phage preparations to GMP and the development of an appropriate regulatory framework that would allow the selection and adaptation of single- or combination phages to be utilised in a personalised manner.

Improvements to phage phenotypic characterisation through further time-kill assays of four candidate phages Bolond, Oda, SteelHaze, and Eggie in combination with antibiotics would greatly improve the understanding of their infection biology. Indeed, further work is necessary to understanding the stability of the phages together in various combinations, particularly with antibiotics, which would provide data on their suitability for phage therapy. Currently, the combination of the three has only been tested against a single *K. michiganensis* strain, and improving the host range capabilities through co-evolution assays could improve their clinical suitability.

Little is known about SteelHaze as it represents a novel genus, and further work to identify its packaging mechanism would also provide a greater understanding of its infection biology, which again, is essential to its use in phage therapy. Future work should also include generating complete, accurate, and high-quality genomes of the remaining draft genomes. In particular, for phage Fifoon, structural predictions with AlphaFold for example, could be used to identify any active site mutations of the DNAP, which would give an indication of whether the phage represents a quasispecies.

Of particular importance would be to generate a highly dense TraDIS mutant library to identify and/or confirm the receptors utilised by each of the candidate phages. A massively-parallel functional genomics screen of the candidate phages with a dense library could reveal insights into their reliance on capsule, their susceptibility genes, and the type of resistance that may emerge.

Together, these data would provide the necessary information to inform clinical suitability and application of the candidate phages against *Klebsiella* infections.

9. REFERENCES

Abedon ST. (2008). Phages, ecology, evolution. *Bacteriophage Ecology: Population Growth, Evolution, and Impact of Bacterial Viruses,* (Abedon ST, ed.) pp. 1-27. Cambridge University Press, Cambridge.

Abedon ST. (2011). Lysis from without. *Bacteriophage* **1**: 46-49.

Abedon ST. (2017). Information phage therapy research should report. *Pharmaceuticals (Basel)* **10**: E43.

Abedon ST. (2019). Phage-antibiotic combination treatments: antagonistic impacts of antibiotics on the pharmacodynamics of phage therapy? *Antibiotics (Basel)* **8**: E182.

Abedon ST, Danis-Wlodarczyk KM and Wozniak DJ. (2021). Phage cocktail development for bacteriophage therapy: toward improving spectrum of activity breadth and depth. *Pharmaceuticals* **14**: 1019.

Ackermann HW. (2005). Bacteriophage classification. *Bacteriophages: Biology and Applications,* (Kutter E & Sulakvelidze A, eds.), pp. 67-89. CRC Press, Boca Raton, Florida.

Acton L, Pye Hannah V, Thilliez G*, et al.* (2024). Collateral sensitivity increases the efficacy of a rationally designed bacteriophage combination to control *Salmonella enterica*. *J Virol* **0**: e01476-01423.

Adams MH. (1959). *Bacteriophages*. Interscience Publishers, New York, New York.

Adewale BA. (2020). Will long-read sequencing technologies replace short-read sequencing technologies in the next 10 years? *Afr J Lab Med* **9**: 1340.

Adler BA, Kazakov AE, Zhong C*, et al.* (2021). The genetic basis of phage susceptibility, cross-resistance and host-range in *Salmonella*. *Microbiology* **167**.

Adriaenssens EM, Ceyssens PJ, Dunon V, Ackermann HW, Van Vaerenbergh J, Maes M, De Proft M and Lavigne R. (2011). Bacteriophages LIMElight and LIMEzero of *Pantoea agglomerans*, belonging to the "phiKMV-like viruses". *Appl Environ Microbiol* **77**: 3443-3450.

Adriaenssens EM, Lehman SM, Vandersteegen K*, et al.* (2012). CIM® monolithic anion-exchange chromatography as a useful alternative to CsCl gradient purification of bacteriophage particles. *Virology* **434**: 265-270.

Adriaenssens EM, Wittmann J, Kuhn JH*, et al.* (2018). Taxonomy of prokaryotic viruses: 2017 update from the ICTV Bacterial and Archaeal Viruses Subcommittee. *Arch Virol* **163**: 1125-1129.

Adriaenssens EM, Sullivan MB, Knezevic P*, et al.* (2020). Taxonomy of prokaryotic viruses: 2018-2019 update from the ICTV Bacterial and Archaeal Viruses Subcommittee. *Arch Virol* **165**: 1253-1260.

Akturk E, Oliveira H, Santos SB, Costa S, Kuyumcu S, Melo LDR and Azeredo J. (2019). Synergistic action of phage and antibiotics: parameters to enhance the killing efficacy against mono and dual-species biofilms. Vol. 8 pp. 103.

Al-Shayeb B, Sachdeva R, Chen LX*, et al.* (2020). Clades of huge phages from across Earth's ecosystems. *Nature* **578**: 425-431.

Alam TI, Draper B, Kondabagil K, Rentas FJ, Ghosh-Kumar M, Sun S, Rossmann MG and Rao VB. (2008). The headful packaging nuclease of bacteriophage T4. *Mol Microbiol* **69**: 1180-1190.

Alavidze Z, Aminov R, Betts A and therapy Ertoaar-iob. (2016). Silk route to the acceptance and re-implementation of bacteriophage therapy. *Biotechnol J* **11**: 595-600.

Alcon-Giner C, Dalby MJ, Caim S*, et al.* (2020). Microbiota supplementation with *Bifidobacterium* and *Lactobacillus* modifies the preterm infant gut microbiota and metabolome: an observational study. *Cell Rep Med* **1**: 100077.

Aleshkin AV, Ershova ON, Volozhantsev NV*, et al.* (2016). Phagebiotics in treatment and prophylaxis of healthcare-associated infections. *Bacteriophage* **6**: e1251379.

Alexander EM, Kreitler DF, Guidolin V, Hurben AK, Drake E, Villalta PW, Balbo S, Gulick AM and Aldrich CC. (2020). Biosynthesis, mechanism of action, and inhibition of the enterotoxin tilimycin produced by the opportunistic pathogen *Klebsiella oxytoca*. *ACS Infect Dis* **6**: 1976-1997.

Alexander MK, Miu A, Oh A, *et al.* (2018). Disrupting Gram-negative bacterial outer membrane biosynthesis through inhibition of the lipopolysaccharide transporter MsbA. *Antimicrob Agents Chemother* **62**: e01142-01118.

Amarasinghe SL, Su S, Dong X, Zappia L, Ritchie ME and Gouil Q. (2020). Opportunities and challenges in long-read sequencing data analysis. *Genome Biol* **21**: 30.

Andrews S. (2010). FastQC: a quality control tool for high throughput sequence data. Available from: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/. (Accessed: 08 Jan 2021).

Anes J, Dever K, Eshwar A, *et al.* (2021). Analysis of the oxidative stress regulon identifies *soxS* as a genetic target for resistance reversal in multidrug-resistant *Klebsiella pneumoniae*. *mBio* **12**: e0086721.

Appel TM, Quijano-Martínez N, De La Cadena E, Mojica MF and Villegas MV. (2021). Microbiological and clinical aspects of *Raoultella* spp. *Front Public Health* **9**: 686789.

Arahal DR. (2014). Whole-genome analyses: average nucleotide identity. *Methods in Microbiology,* Vol. 41 (Goodfellow M, Sutcliffe I & Chun J, eds.), pp. 103-122. Academic Press, London.

Arizala D and Arif M. (2019). Genome-wide analyses revealed remarkable heterogeneity in pathogenicity determinants, antimicrobial compounds, and CRISPR-Cas systems of complex phytopathogenic genus *Pectobacterium*. *Pathogens* **8**.

Azeredo J and Sutherland IW. (2008). The use of phages for the removal of infectious biofilms. *Curr Pharm Biotechnol* **9**: 261-266.

Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, Datsenko KA, Tomita M, Wanner BL and Mori H. (2006). Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol* **2**: 2006.0008.

Bailly-Bechet M, Vergassola M and Rocha E. (2007). Causes for the intriguing presence of tRNAs in phages. *Genome Res* **17**: 1486-1495.

Bansal S, Harjai K and Chhibber S. (2014). Depolymerase improves gentamicin efficacy during *Klebsiella pneumoniae* induced murine infection. *BMC Infect Dis* **14**: 456.

Bansal V and Boucher C. (2019). Sequencing Technologies and Analyses: Where Have We Been and Where Are We Going? *iScience* **18**: 37-41.

Baptista C, Santos MA and São-José C. (2008). Phage SPP1 reversible adsorption to *Bacillus subtilis* cell wall teichoic acids accelerates virus recognition of membrane receptor YueB. *J Bacteriol* **190**: 4989-4996.

Barquist L, Mayho M, Cummins C, Cain AK, Boinett CJ, Page AJ, Langridge GC, Quail MA, Keane JA and Parkhill J. (2016). The TraDIS toolkit: sequencing and analysis for dense transposon mutant libraries. *Bioinformatics* **32**: 1109-1111.

Barylski J, Kropinski AM, Alikhan NF, Adriaenssens EM and ICTV Report Consortium. (2020). ICTV virus taxonomy profile: *Herelleviridae*. *J Gen Virol*.

Barylski J, Enault F, Dutilh BE, *et al.* (2020). Analysis of Spounaviruses as a case study for the overdue reclassification of tailed phages. *Syst Biol* **69**: 110-123.

Basak S, Singh P and Rajurkar M. (2016). Multidrug resistant and extensively drug resistant bacteria: a study. *J Pathog* **2016**: 4065603.

Beamud B, García-González N, Gómez-Ortega M, González-Candelas F, Domingo-Calap P and Sanjuan R. (2023). Genetic determinants of host tropism in *Klebsiella* phages. *Cell Rep* **42**: 112048.

Benes V, Kilger C, Voss H, Pääbo S and Ansorge W. (1997). Direct primer walking on P1 plasmid DNA. *Biotechniques* **23**: 98-100.

Bengoechea JA and Sa Pessoa J. (2019). *Klebsiella pneumoniae* infection biology: living to counteract host defences. *FEMS Microbiol Rev* **43**: 123-144.

Berry GJ, Loeffelholz MJ and Williams-Bouyer N. (2015). An Investigation into laboratory misidentification of a bloodstream *Klebsiella variicola* Infection. *J Clin Microbiol* **53**: 2793-2794.

Berry J, Summer EJ, Struck DK and Young R. (2008). The final step in the phage infection cycle: the Rz and Rz1 lysis proteins link the inner and outer membranes. *Mol Microbiol* **70**: 341-351.

Bertozzi Silva J, Storms Z and Sauvageau D. (2016). Host receptors for bacteriophage adsorption. *FEMS Microbiol Lett* **363**: fnw002.

Bharti R and Grimm DG. (2021). Current challenges and best-practice protocols for microbiome analysis. *Brief Bioinform* **22**: 178-193.

Bird JT, Burke KA, Urick CD, Braverman JL, Mzhavia N, Ellison DW, Nikolich MP and Filippov AA. (2024). Genome sequence of the *Klebsiella quasipneumoniae* bacteriophage EKq1 with activity against *Klebsiella pneumoniae. Microbiol Resour Announc* **13**: e0095423.

Black LW. (1989). DNA packaging in dsDNA bacteriophages. *Annu Rev Microbiol* **43**: 267-292.

Blair JM, Webber MA, Baylay AJ, Ogbolu DO and Piddock LJ. (2015). Molecular mechanisms of antibiotic resistance. *Nat Rev Microbiol* **13**: 42-51.

Bleriot I, Blasco L, Pacios O*, et al.* (2023). Proteomic study of the interactions between phages and the bacterial host *Klebsiella pneumoniae. Microbiol Spectr* **11**: e0397422.

Bleriot I, Blasco L, Pacios O*, et al.* (2022). The role of PemIK (PemK/PemI) type II TA system from *Klebsiella pneumoniae* clinical strains in lytic phage infection. *Sci Rep* **12**: 4488.

Boeckman J, Korn A, Yao G, Ravindran A, Gonzalez C and Gill J. (2022). Sheep in wolves' clothing: Temperate T7-like bacteriophages and the origins of the Autographiviridae. *Virology* **568**: 86-100.

Bollenbach T. (2015). Antimicrobial interactions: mechanisms and implications for drug discovery and resistance evolution. *Curr Opin Microbiol* **27**: 1-9.

Bonilla N, Rojas MI, Netto Flores Cruz G, Hung S-H, Rohwer F and Barr JJ. (2016). Phage on tap–a quick and efficient protocol for the preparation of bacteriophage laboratory stocks. *PeerJ* **4**: e2261.

Borodovich T, Shkoporov AN, Ross RP and Hill C. (2022). Phage-mediated horizontal gene transfer and its implications for the human gut microbiome. *Gastroenterol Rep* **10**: goac012.

Bortolaia V, Kaas RS, Ruppe E*, et al.* (2020). ResFinder 4.0 for predictions of phenotypes from genotypes. *J Antimicrob Chemother* **75**: 3491-3500.

Bosco K, Lynch S, Sandaradura I and Khatami A. (2023). Therapeutic phage monitoring: a review. *Clin Infect Dis* **77**: S384-S394.

Bouras G, Nepal R, Houtak G, Psaltis AJ, Wormald P-J and Vreugde S. (2022). Pharokka: a fast scalable bacteriophage annotation tool. *Bioinformatics* **39**: btac776.

Bourque G, Burns KH, Gehring M*, et al.* (2018). Ten things you should know about transposable elements. *Genome Biol* **19**: 199.

Bowers JR, Kitchel B, Driebe EM*, et al.* (2015). Genomic analysis of the emergence and rapid global dissemination of the clonal group 258 *Klebsiella pneumoniae* pandemic. *PLoS One* **10**: e0133727.

Brisse S and Verhoef J. (2001). Phylogenetic diversity of *Klebsiella pneumoniae* and *Klebsiella oxytoca* clinical isolates revealed by randomly amplified polymorphic DNA, *gyrA* and *parC* genes sequencing and automated ribotyping. *Int J Syst Evol Microbiol* **51**: 915-924.

Brisse S, Passet V and Grimont PA. (2014). Description of *Klebsiella quasipneumoniae* sp. nov., isolated from human infections, with two subspecies, *Klebsiella quasipneumoniae* subsp. *quasipneumoniae* subsp. nov. and *Klebsiella quasipneumoniae* subsp. *similipneumoniae* subsp. nov., and demonstration that *Klebsiella singaporensis* is a junior heterotypic synonym of *Klebsiella variicola*. *Int J Syst Evol Microbiol* **64**: 3146-3152.

Broberg CA, Palacios M and Miller VL. (2014). *Klebsiella*: a long way to go towards understanding this enigmatic jet-setter. *F1000Prime Rep* **6**: 64.

Brown TL, Petrovski S, Hoyle D, Chan HT, Lock P and Tucci J. (2017). Characterization and formulation into solid dosage forms of a novel bacteriophage lytic against *Klebsiella oxytoca. PLoS One* **12**: e0183510.

Bruchmann S, Feltwell T, Parkhill J and Short FL. (2021). Identifying virulence determinants of multidrug-resistant *Klebsiella pneumoniae* in *Galleria mellonella. Pathog Dis* **79**.

Brüssow H and Kutter E. (2005). Genomics and evolution of tailed phages. *Bacteriophages: Biology and Applications,* (Kutter E & Sulakvelidze A, eds.), pp. 91-128. CRC Press, Boca Raton, Florida.

Buckner MMC, Saw HTH, Osagie RN, McNally A, Ricci V, Wand ME, Woodford N, Ivens A, Webber MA and Piddock LJV. (2018). Clinically relevant plasmid-host interactions indicate that transcriptional and not genomic modifications ameliorate fitness costs of *Klebsiella pneumoniae* carbapenemase-carrying plasmids. *mBio* **9**: e02303-02317.

Buffet A, Rocha EPC and Rendueles O. (2021). Nutrient conditions are primary drivers of bacterial capsule maintenance in *Klebsiella*. *Proc Biol Sci* **288**: 20202876.

Bulssico J, Papukashvill I, Espinosa L, Gandon S and Ansaldi M. (2023). Phage-antibiotic synergy: cell filamentation is a key driver of successful phage predation. *PLoS Pathogens* **19**: e1011602.

Byarugaba DK, Erima B, Wokorach G, Alafi S, Kibuuka H, Mworozi E, Najjuka F, Kiyengo J, Musinguzi AK and Wabwire-Mangen F. (2023). Genome analysis of *Klebsiella pneumoniae* reveals international high-risk pandemic MDR clones emerging in tertiary healthcare settings in Uganda. *Pathogens* **12**.

Cai R, Wu M, Zhang H*, et al.* (2018). A smooth-type, phage-resistant *Klebsiella pneumoniae* mutant strain reveals that OmpC is indispensable for infection by phage GH-K3. *Appl Environ Microbiol* **84**: e01585-01518.

Cai R, Wang Z, Wang G*, et al.* (2019). Biological properties and genomics analysis of vB_KpnS_GH-K3, a *Klebsiella* phage with a putative depolymerase-like protein. *Virus Genes* **55**: 696-706.

Cain AK, Barquist L, Goodman AL, Paulsen IT, Parkhill J and van Opijnen T. (2020). A decade of advances in transposon-insertion sequencing. *Nat Rev Genet*.

Callanan J, Stockdale SR, Shkoporov A, Draper LA, Ross RP and Hill C. (2018). RNA phage biology in a metagenomic era. *Viruses* **10**: E386.

Callanan J, Stockdale SR, Adriaenssens EM, Kuhn JH, Rumnieks J, Pallen MJ, Shkoporov AN, Draper LA, Ross RP and Hill C. (2021). *Leviviricetes*: expanding and restructuring the taxonomy of bacteria-infecting single-stranded RNA viruses. *Microb Genom* **7**.

Camargo AP, Roux S, Schulz F, Babinski M, Xu Y, Hu B, Chain PSG, Nayfach S and Kyrpides NC. (2023). Identification of mobile genetic elements with geNomad. *Nat Biotechnol*.

Cano EJ, Caflisch KM, Bollyky PL*, et al.* (2021). Phage therapy for limb-threatening prosthetic knee *Klebsiella pneumoniae* infection: case report and in vitro characterization of anti-biofilm activity. *Clin Infect Dis* **73**: e144-e151.

Cantón R, González-Alba JM and Galán JC. (2012). CTX-M enzymes: origin and diffusion. *Front Microbiol* **3**: 110.

Carabarin-Lima A, Leon-Izurieta L, Rocha-Gracia RDC, Castaneda-Lucio M, Torres C, Gutierrez-Cazarez Z, Gonzalez-Posos S, Martinez de la Pena CF, Martinez-Laguna Y and Lozano-Zarain P. (2016). First evidence of polar flagella in *Klebsiella pneumoniae* isolated from a patient with neonatal sepsis. *J Med Microbiol* **65**: 729-737.

Carattoli A, Zankari E, Garcia-Fernandez A, Voldby Larsen M, Lund O, Villa L, Moller Aarestrup F and Hasman H. (2014). *In silico* detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. *Antimicrob Agents Chemother* **58**: 3895-3903.

Carpinelli J, Krämer R and Agosin E. (2006). Metabolic engineering of *Corynebacterium glutamicum* for trehalose overproduction: role of the TreYZ trehalose biosynthetic pathway. *Appl Environ Microbiol* **72**: 1949-1955.

Carver T, Thomson N, Bleasby A, Berriman M and Parkhill J. (2008). DNAPlotter: circular and linear interactive genome visualization. *Bioinformatics* **25**: 119-120.

Carver T, Harris SR, Berriman M, Parkhill J and McQuillan JA. (2012). Artemis: an integrated platform for visualization and analysis of high-throughput sequence-based experimental data. *Bioinformatics* **28**: 464-469.

Casjens SR and Gilcrease EB. (2009). Determining DNA packaging strategy by analysis of the termini of the chromosomes in tailed-bacteriophage virions. *Methods Mol Biol* **502**: 91-111.

Cassini A, Högberg LD, Plachouras D*, et al.* (2019). Attributable deaths and disability-adjusted life-years caused by infections with antibiotic-resistant bacteria in the EU and the European Economic Area in 2015: a population-level modelling analysis. *Lancet Infect Dis* **19**: 56-66.

Celia F-G, Robby C-E, Mireia B-G, Felipe F-C, Javier EC-G, Silvia G-C, Rafael S and Pilar D-C. (2024). Targeted phage hunting to specific *Klebsiella pneumoniae* clinical isolates is an efficient antibiotic resistance and infection control strategy. *bioRxiv*.

Cenens W, Makumi A, Mebrhatu MT, Lavigne R and Aertsen A. (2013). Phage-host interactions during pseudolysogeny: lessons from the Pid/*dgo* interaction. *Bacteriophage* **3**: e25029.

Centre for Genomic Pathogen Surveillance. (2024). Pathogen Watch. Available from: https://pathogen.watch/. (Accessed: 12 June 2024).

Chagneau CV, Payros D, Tang-Fichaux M, Auvray F, Nougayrède J-P and Oswald E. (2022). The pks island: a bacterial Swiss army knife? Colibactin: beyond DNA damage and cancer. *Trends in Microbiology* **30**: 1146-1159.

Chan BK, Sistrom M, Wertz JE, Kortright KE, Narayan D and Turner PE. (2016). Phage selection restores antibiotic sensitivity in MDR *Pseudomonas aeruginosa*. *Sci Rep* **6**: 26717.

Chatterjee S and Rothenberg E. (2012). Interaction of bacteriophage λ with its *E. coli* receptor, LamB. *Viruses* **4**: 3162-3178.

Chen L, Mathema B, Pitout JD, DeLeo FR and Kreiswirth BN. (2014). Epidemic *Klebsiella pneumoniae* ST258 is a hybrid strain. *mBio* **5**: e01355-01314.

Chen M, Li Y, Li S, Tang L, Zheng J and An Q. (2016). Genomic identification of nitrogen-fixing *Klebsiella variicola*, *K. pneumoniae* and *K. quasipneumoniae*. *J Basic Microbiol* **56**: 78-84.

Chen S, Zhou Y, Chen Y and Gu J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**: i884-i890.

Chen Y, Zhang Y, Wang AY, Gao M and Chong Z. (2021). Accurate long-read *de novo* assembly evaluation with Inspector. *Genome Biology* **22**: 312.

Chen Y, Brook TC, Soe CZ*, et al.* (2020). Preterm infants harbour diverse *Klebsiella* populations, including atypical species that encode and produce an array of antimicrobial resistance- and virulence-associated factors. *Microb Genom*.

Chernyatina AA and Low HH. (2019). Core architecture of a bacterial type II secretion system. *Nature Communications* **10**: 5437.

Chevallereau A, Pons BJ, van Houte S and Westra ER. (2022). Interactions between bacterial and phage communities in natural environments. *Nat Rev Microbiol* **20**: 49-62.

Chevallereau A, Meaden S, Fradet O, Landsberger M, Maestri A, Biswas A, Gandon S, van Houte S and Westra ER. (2020). Exploitation of the cooperative behaviors of anti-CRISPR phages. *Cell Host Microbe* **27**: 189-198.

Chhibber S, Nag D and Bansal S. (2013). Inhibiting biofilm formation by *Klebsiella pneumoniae* B5055 using an iron antagonizing molecule and a bacteriophage. *BMC Microbiol* **13**: 174.

Chiang YN, Penadés JR and Chen J. (2019). Genetic transduction by phages and chromosomal islands: the new and noncanonical. *PLoS Pathog* **15**: e1007878.

Choby JE, Howard-Anderson J and Weiss DS. (2020). Hypervirulent *Klebsiella pneumoniae* – clinical and molecular perspectives. *J Intern Med* **287**: 283-300.

Choi KH. (2012). Viral Polymerases. *Viral Molecular Machines,* (Rossmann MG & Rao VB, eds.), pp. 267-304. Springer US, Boston, MA.

Clokie MR, Millard AD, Letarov AV and Heaphy S. (2011). Phages in nature. *Bacteriophage* **1**: 31-45.

Comeau AM, Tétart F, Trojet SN, Prère MF and Krisch HM. (2007). Phage-antibiotic synergy (PAS): β-lactam and quinolone antibiotics stimulate virulent phage growth. *PLoS One* **2**: e799.

Cook R, Brown N, Rihtman B*, et al.* (2024). The long and short of it: benchmarking viromics using Illumina, Nanopore and PacBio sequencing technologies. *Microb Genom* **10**.

Corbellino M, Kieffer N, Kutateladze M*, et al.* (2020). Eradication of a multidrug-resistant, carbapenemase-producing *Klebsiella pneumoniae* isolate following oral and intra-rectal therapy with a custom made, lytic bacteriophage preparation. *Clin Infect Dis* **70**: 1998-2001.

Corelli B, Almeida AS, Sonego F, Castiglia V, Fevre C, Brisse S, Sansonetti PJ and Tournebize R. (2018). Rhinoscleroma pathogenesis: the type K3 capsule of *Klebsiella rhinoscleromatis* is a virulence factor not involved in Mikulicz cells formation. *PLoS Negl Trop Dis* **12**: e0006201-e0006201.

Cowley LA, Low AS, Pickard D*, et al.* (2018). Transposon insertion sequencing elucidates novel gene involvement in susceptibility and resistance to phages T4 and T7 in *Escherichia coli* O157. *mBio* **9**.

Crippa C, Pasquali F, Rodrigues C, De Cesare A, Lucchi A, Gambi L, Manfreda G, Brisse S and Palma F. (2023). Genomic features of *Klebsiella* isolates from artisanal ready-to-eat food production facilities. *Sci Rep* **13**: 10957.

Cui Z, Shen W, Wang Z*, et al.* (2012). Complete genome sequence of *Klebsiella pneumoniae* phage JD001. *J Virol* **86**: 13843.

d'Hérelle F. (1917). Sur un microbe invisible antagoniste des bacilles dysentériques. *CR Acad Sci Paris* **165**: 373-375.

d'Hérelle F and Smith GH. (1926). The bacteriophage and its behaviour. *Nature* **118**: 183-185.

Darby A, Lertpiriyapong K, Sarkar U*, et al.* (2014). Cytotoxic and pathogenic properties of *Klebsiella oxytoca* isolated from laboratory animals. *PLoS One* **9**: e100542.

Darling AE, Mau B and Perna NT. (2010). progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS One* **5**: e11147.

Davies EV, Winstanley C, Fothergill JL and James CE. (2016). The role of temperate bacteriophages in bacterial infection. *FEMS Microbiol Lett* **363**: fnw015.

De Coster W, D'Hert S, Schultz DT, Cruts M and Van Broeckhoven C. (2018). NanoPack: visualizing and processing long-read sequencing data. *Bioinformatics* **34**: 2666-2669.

Debarbieux L, Bohin A and Bohin JP. (1997). Topological analysis of the membrane-bound glucosyltransferase, MdoH, required for osmoregulated periplasmic glucan synthesis in *Escherichia coli*. *J Bacteriol* **179**: 6692-6698.

Decraene V, Phan HTT, George R*, et al.* (2018). A large, refractory nosocomial outbreak of *Klebsiella pneumoniae* carbapenemase-producing *Escherichia coli* demonstrates carbapenemase gene outbreaks involving sink sites require novel approaches to infection control. *Antimicrob Agents Chemother* **62**: e01689-01618.

Dedrick RM, Guerrero-Bustamante CA, Garlena RA*, et al.* (2019). Engineered bacteriophages for treatment of a patient with a disseminated drug-resistant *Mycobacterium abscessus*. *Nat Med* **25**: 730-733.

Diancourt L, Passet V, Verhoef J, Grimont PA and Brisse S. (2005). Multilocus sequence typing of *Klebsiella pneumoniae* nosocomial isolates. *J Clin Microbiol* **43**: 4178-4182.

Díaz E, Ferrández A and García JL. (1998). Characterization of the *hca* cluster encoding the dioxygenolytic pathway for initial catabolism of 3-phenylpropionic acid in *Escherichia coli* K-12. *J Bacteriol* **180**: 2915-2923.

Dimitriu T, Kurilovich E, Łapińska U, Severinov K, Pagliara S, Szczelkun MD and Westra ER. (2022). Bacteriostatic antibiotics promote CRISPR-Cas adaptive immunity by enabling increased spacer acquisition. *Cell Host Microbe* **30**: 31-40 e35.

Dion MB, Oechslin F and Moineau S. (2020). Phage diversity, genomics and phylogeny. *Nat Rev Microbiol* **18**: 125-138.

Domingo E and Perales C. (2019). Viral quasispecies. *PLoS Genet* **15**: e1008271.

Donker T, Henderson KL, Hopkins KL*, et al.* (2017). The relative importance of large problems far away versus small problems closer to home: insights into limiting the spread of antimicrobial resistance in England. *BMC Med* **15**: 86.

Dufour N, Delattre R, Ricard JD and Debarbieux L. (2017). The lysis of pathogenic *Escherichia coli* by bacteriophages releases less endotoxin than by β-lactams. *Clin Infect Dis* **64**: 1582-1588.

Dufour N, Delattre R, Chevallereau A, Ricard JD and Debarbieux L. (2019). Phage therapy of pneumonia is not associated with an overstimulation of the inflammatory response compared to antibiotic treatment in mice. *Antimicrob Agents Chemother* **63**: e00379-00319.

Dunn JJ and Studier FW. (1983). Complete nucleotide sequence of bacteriophage T7 DNA and the locations of T7 genetic elements. *J Mol Biol* **166**: 477-535.

Duo M, Hou S and Ren D. (2008). Identifying *Escherichia coli* genes involved in intrinsic multidrug resistance. *Appl Microbiol Biotechnol* **81**: 731-741.

Edgar R, Rokney A, Feeney M, Semsey S, Kessel M, Goldberg MB, Adhya S and Oppenheim AB. (2008). Bacteriophage infection is targeted to cellular poles. *Mol Microbiol* **68**: 1107-1116.

Edgar RC. (2004). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinform* **5**: 113.

Elek CKA, Brown TL, Le Viet T*, et al.* (2023). A hybrid and poly-polish workflow for the complete and accurate assembly of phage genomes: a case study of ten przondoviruses. *Microb Genom* **9**: mgen001065.

ESCMID. (2000). In EUCAST Definitive Document E.Def 1.2: Terminology relating to methods for the determination of susceptibility of bacteria to antimicrobial agents. *Clin Microbiol Infect* **6**: 503-508.

Eskenazi A, Lood C, Wubbolts J*, et al.* (2022). Combination of pre-adapted bacteriophage therapy and antibiotics for treatment of fracture-related infection due to pandrug-resistant Klebsiella pneumoniae. *Nat Commun* **13**: 302.

EUCAST. (2000). EUCAST Definitive Document E.DEF 2.1, August 2000: Determination of antimicrobial susceptibility test breakpoints. *Clin Microbiol Infect* **6**: 570-572.

EUCAST. (2023). Clinical Breakpoints and Dosing of Antibiotics. Available from: https://www.eucast.org/clinical_breakpoints. (Accessed: 24 Jul 2023).

European Commission. (2001). Directive 2001/83/EC of the European Parliament and of the Council of 6 November 2001 on the Community Code Relating to Medicinal Products for Human Use. (Commission E, ed.) pp. European Commission, Brussels.

Evseev PV, Lukianova AA, Shneider MM*, et al.* (2020). Origin and evolution of *Studiervirinae* bacteriophages infecting *Pectobacterium*: horizontal transfer assists adaptation to new niches. *Microorganisms* **8**.

Fabianek RA, Hofer T and Thöny-Meyer L. (1999). Characterization of the *Escherichia coli* CcmH protein reveals new insights into the redox pathway required for cytochrome c maturation. *Arch Microbiol* **171**: 92-100.

Fane BA, Brentlinger KL, Burch AD, Chen M, Hafenstein S, Moore E, Novak CR and Uchiyama A. (2006). φX174 et al., the *Microviridae*. *The Bacteriophages,* (Calendar R, ed.) pp. 129-145. Oxford University Press, Oxford.

Fang C, Dai X, Xiang L, Qiu Y, Yin M, Fu Y, Li Y and Zhang L. (2023). Isolation and characterization of three novel lytic phages against K54 serotype carbapenem-resistant hypervirulent *Klebsiella pneumoniae*. *Front Cell Infect Microbiol* **13**: 1265011.

Federici S, Kredo-Russo S, Valdés-Mas R*, et al.* (2022). Targeted suppression of human IBD-associated gut microbiota commensals by phage consortia for treatment of intestinal inflammation. *Cell* **185**: 2879-2898.e2824.

Feil EJ and Spratt BG. (2001). Recombination and the population structures of bacterial pathogens. *Annu Rev Microbiol* **55**: 561-590.

Fevre C, Passet V, Weill FX, Grimont PAD and Brisse S. (2005). Variants of the *Klebsiella pneumoniae* OKP chromosomal β-lactamase are divided into two main groups, OKP-A and OKP-B. *Antimicrob Agents Chemother* **49**: 5149-5152.

Fevre C, Jbel M, Passet V, Weill FX, Grimont PA and Brisse S. (2005). Six groups of the OXY β-lactamase evolved over millions of years in *Klebsiella oxytoca*. *Antimicrob Agents Chemother* **49**: 3453-3462.

Filenko N, Spiro S, Browning Douglas F, Squire D, Overton Tim W, Cole J and Constantinidou C. (2007). The NsrR regulon of *Escherichia coli* K-12 includes genes encoding the hybrid cluster protein and the periplasmic, respiratory nitrite reductase. *J Bacteriol* **189**: 4410-4417.

Florensa AF, Kaas RS, Clausen P, Aytan-Aktug D and Aarestrup FM. (2022). ResFinder - an open online resource for identification of antimicrobial resistance genes in next-generation sequencing data and prediction of phenotypes from genotypes. *Microb Genom* **8**: 000748.

Fontana L, Bonura E, Lyski Z and Messer W. (2019). Closing the brief case: *Klebsiella variicola* - identifying the misidentified. *J Clin Microbiol* **57**: e00825-00818.

Friman VP, Soanes-Brown D, Sierocinski P, Molin S, Johansen HK, Merabishvili M, Pirnay JP, De Vos D and Buckling A. (2016). Pre-adapting parasitic phages to a pathogen leads to increased pathogen clearance and lowered resistance evolution with *Pseudomonas aeruginosa* cystic fibrosis bacterial isolates. *J Evol Biol* **29**: 188-198.

Fung C-P, Hu B-S, Chang F-Y, Lee S-C, In-Tiau Kuo B, Ho M, Siu LK and Liu C-Y. (2000). A 5-year study of the seroepidemiology of *Klebsiella pneumoniae*: high prevalence of

capsular serotype K1 in Taiwan and implication for vaccine efficacy. *J Infect Dis* **181**: 2075-2079.

Fung CP, Chang FY, Lin JC, Ho DMT, Chen CT, Chen JH, Yeh KM, Chen TL, Lin YT and Siu LK. (2011). Immune response and pathophysiological features of *Klebsiella pneumoniae* liver abscesses in an animal model. *Lab Invest* **91**: 1029-1039.

Furfaro LL, Payne MS and Chang BJ. (2018). Bacteriophage therapy: clinical trials and regulatory hurdles. *Front Cell Infect Microbiol* **8**: 376.

Gao D, Ji H, Wang L, Li X, Hu D, Zhao J, Wang S, Tao P, Li X and Qian P. (2022). Fitness trade-offs in phage cocktail-resistant *Salmonella enterica* serovar Enteritidis results in increased antibiotic susceptibility and reduced virulence. *Microbiol Spectr* **10**: e0291422.

Gao M, Yi L, Wang Y, Gao J, Liu H, Zhang X, Pei G, Tong Y and Bai C. (2022). Characterization and genomic analysis of bacteriophage vB_KpnM_IME346 targeting clinical *Klebsiella pneumoniae* strain of the K63 capsular type. *Curr Microbiol* **79**: 160.

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

Geier MR, Trigg ME and Merril CR. (1973). Fate of bacteriophage lambda in non-immune germ-free mice. *Nature* **246**: 221-223.

Giani AM, Gallo GR, Gianfranceschi L and Formenti G. (2020). Long walk to genomics: history and current approaches to genome sequencing and assembly. *Comput Struct Biotechnol J* **18**: 9-19.

Gibson MK, Wang B, Ahmadi S, Burnham CAD, Tarr PI, Warner BB and Dantas G. (2016). Developmental dynamics of the preterm infant gut microbiota and antibiotic resistome. *Nat Microbiol* **1**: 16024.

Gibson SB, Green SI, Liu CG, Salazar KC, Clark JR, Terwilliger AL, Kaplan HB, Maresso AW, Trautner BW and Ramig RF. (2019). Constructing and characterizing bacteriophage libraries for phage therapy of human infections. *Front Microbiol* **10**: 2537.

Gilchrist CLM and Chooi Y-H. (2021). Clinker & clustermap.js: automatic generation of gene cluster comparison figures. *Bioinformatics* **37**: 2473-2475.

González-García VA, Pulido-Cid M, Garcia-Doval C, Bocanegra R, van Raaij MJ, Martín-Benito J, Cuervo A and Carrascosa JL. (2015). Conformational changes leading to T7 DNA delivery upon interaction with the bacterial receptor. *J Biol Chem* **290**: 10038-10044.

González-López JJ, Coelho A, Larrosa MN, Lavilla S, Bartolomé R and Prats G. (2009). First detection of plasmid-encoded bla_{OXY} β-lactamase. *Antimicrob Agents Chemother* **53**: 3143-3146.

Gonzalez-Serrano R, Rosselli R, Roda-Garcia JJ, Martin-Cuadrado AB, Rodriguez-Valera F and Dunne M. (2023). Distantly related *Alteromonas* bacteriophages share tail fibers exhibiting properties of transient chaperone caps. *Nat Commun* **14**: 6517.

Gorbalenya AE, Lauber C and Siddell S. (2019). Taxonomy of Viruses. *Reference Module in Biomedical Sciences,* pp. Elsevier.

Gorbet MB and Sefton MV. (2005). Endotoxin: the uninvited guest. *Biomaterials* **26**: 6811-6817.

Gorrie CL, Mirceta M, Wick RR*, et al.* (2018). Antimicrobial-resistant *Klebsiella pneumoniae* carriage and infection in specialized geriatric care wards linked to acquisition in the referring hospital. *Clin Infect Dis* **67**: 161-170.

Gramer J, Kenny S, Newkirk H, Liu M, Gill JJ and Ramsey J. (2019). Complete genome sequence of *Klebsiella pneumoniae* siphophage Skenny. *Microbiol Resour Announc* **8**: e01036-01019.

Grasis JA. (2018). Host-associated bacteriophage isolation and preparation for viral metagenomics. *Viral Metagenomics: Methods and Protocols,* (Pantaleo V & Chiumenti M, eds.), pp. 1-25. Springer New York, New York, New York.

Grenier F, Matteau D, Baby V and Rodrigue S. (2014). Complete genome sequence of *Escherichia coli* BW25113. *Genome Announc* **2**: e01038-01014.

Grigoriev A. (1998). Analyzing genomes with cumulative skew diagrams. *Nucleic Acids Res* **26**: 2286-2290.

Grubaugh ND, Gangavarapu K, Quick J, *et al.* (2019). An amplicon-based sequencing framework for accurately measuring intrahost virus diversity using PrimalSeq and iVar. *Genome Biol* **20**: 8.

Gu Liu C, Green SI, Min L, *et al.* (2020). Phage-antibiotic synergy is driven by a unique combination of antibacterial mechanism of action and stoichiometry. *mBio* **11**.

Gu Z. (2022). Complex heatmap visualization. *iMeta* **1**: e43.

Gu Z, Eils R and Schlesner M. (2016). Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**: 2847-2849.

Guo Q, Spychala CN, McElheny CL and Doi Y. (2016). Comparative analysis of an IncR plasmid carrying *armA*, *bla*$_{DHA-1}$ and *qnrB4* from *Klebsiella pneumoniae* ST37 isolates. *J Antimicrob Chemother* **71**: 882-886.

Guo X, Soderholm A, Kanchugal PS, *et al.* (2021). Structure and mechanism of a phage-encoded SAM lyase revises catalytic function of enzyme family. *Elife* **10**.

Guttman B, Raya R and Kutter E. (2005). Basic phage biology. *Bacteriophages: Biology and Applications,* (Kutter E & Sulakvelidze A, eds.), pp. 29-66. CRC Press, Boca Raton, Florida.

Guzmán K, Badia J, Giménez R, Aguilar J and Baldoma L. (2011). Transcriptional regulation of the gene cluster encoding allantoinase and guanine deaminase in *Klebsiella pneumoniae*. *J Bacteriol* **193**: 2197-2207.

Hadfield J, Croucher NJ, Goater RJ, Abudahab K, Aanensen DM and Harris SR. (2017). Phandango: an interactive viewer for bacterial population genomics. *Bioinformatics* **34**: 292-293.

Hæggman S, Löfdahl S, Paauw A, Verhoef J and Brisse S. (2004). Diversity and evolution of the class A chromosomal beta-lactamase gene in *Klebsiella pneumoniae*. *Antimicrob Agents Chemother* **48**: 2400-2408.

Hamdi S, Rousseau GM, Labrie SJ, Kourda RS, Tremblay DM, Moineau S and Slama KB. (2016). Characterization of five *Podoviridae* phages infecting *Citrobacter freundii*. *Front Microbiol* **7**: 1023.

Harding NE, Cleary JM, Smith DW, Michon JJ, Brusilow WS and Zyskind JW. (1982). Chromosomal replication origins (oriC) of *Enterobacter aerogenes* and *Klebsiella pneumoniae* are functional in *Escherichia coli*. *J Bacteriol* **152**: 983-993.

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

Haudiquet M, Le Bris J, Nucci A, Bonnin RA, Domingo-Calap P, Rocha EPC and Rendueles O. (2024). Capsules and their traits shape phage susceptibility and plasmid conjugation efficiency. *Nat Commun* **15**: 2032.

Hay ID and Lithgow T. (2019). Filamentous phages: masters of a microbial sharing economy. *EMBO Rep* **20**: e47427.

Hendrix RW. (2002). Bacteriophages: evolution of the majority. *Theor Popul Biol* **61**: 471-480.

Hering NA, Fromm A, Bücker R, Gorkiewicz G, Zechner E, Högenauer C, Fromm M, Schulzke J-D and Troeger H. (2019). Tilivalline- and tilimycin-independent effects of *Klebsiella oxytoca* on tight junction-mediated intestinal barrier impairment. *Int J Mol Sci* **20**: E5595.

Herzog KA, Schneditz G, Leitner E, Feierl G, Hoffmann KM, Zollner-Schwetz I, Krause R, Gorkiewicz G, Zechner EL and Högenauer C. (2014). Genotypes of *Klebsiella oxytoca* isolates from patients with nosocomial pneumonia are distinct from those of isolates from patients with antibiotic-associated hemorrhagic colitis. *J Clin Microbiol* **52**: 1607-1616.

Hesse S, Rajaure M, Wall E, Johnson J, Bliskovsky V, Gottesman S and Adhya S. (2020). Phage resistance in multidrug-resistant *Klebsiella pneumoniae* ST258 evolves via diverse mutations that culminate in impaired adsorption. *mBio* **11**: e02530-02519.

Hilbert BJ, Hayes JA, Stone NP, Xu R-G and Kelch BA. (2017). The large terminase DNA packaging motor grips DNA with its ATPase domain for cleavage by the flexible nuclease domain. *Nucleic Acids Res* **45**: 3591-3605.

Hiltunen T, Virta M and Laine AL. (2017). Antibiotic resistance in the wild: an eco-evolutionary perspective. *Philos Trans R Soc Lond B Biol Sci* **372**: 20160039.

Hinton DM. (2010). Transcriptional control in the prereplicative phase of T4 development. *Virol J* **7**: 289.

Hirooka K, Kodoi Y, Satomura T and Fujita Y. (2016). Regulation of the *rhaEWRBMA* operon involved in L-rhamnose catabolism through two transcriptional factors, RhaR and CcpA, in *Bacillus subtilis*. *J Bacteriol* **198**: 830-845.

Hoang DT, Chernomor O, von Haeseler A, Minh BQ and Vinh LS. (2017). UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Mol Biol Evol* **35**: 518-522.

Hodyra-Stefaniak K, Miernikiewicz P, Drapała J*, et al.* (2015). Mammalian host-versus-phage immune response determines phage fate *in vivo*. *Sci Rep* **5**: 14802.

Hoffman LM, Jendrisak JJ, Meis RJ, Goryshin IY and Reznikof SW. (2000). Transposome insertional mutagenesis and direct sequencing of microbial genomes. *Genetica* **108**: 19-24.

Högenauer C, Langner C, Beubler E, Lippe IT, Schicho R, Gorkiewicz G, Krause R, Gerstgrasser N, Krejs GJ and Hinterleitner TA. (2006). *Klebsiella oxytoca* as a causative organism of antibiotic-associated hemorrhagic colitis. *N Engl J Med* **355**: 2418-2426.

Holden ER, Yasir M, Turner AK, Wain J, Charles IG and Webber MA. (2021). Massively parallel transposon mutagenesis identifies temporally essential genes for biofilm formation in Escherichia coli. *Microb Genom* **7**: 000673.

Holden VI, Breen P, Houle S, Dozois CM and Bachman MA. (2016). *Klebsiella pneumoniae* siderophores induce inflammation, bacterial dissemination, and HIF-1α stabilization during pneumonia. *mBio* **7**: e01397-01316.

Holt KE, Lassalle F, Wyres KL, Wick R and Mostowy RJ. (2020). Diversity and evolution of surface polysaccharide synthesis loci in Enterobacteriales. *ISME J*.

Holt KE, Wertheim H, Zadoks RN*, et al.* (2015). Genomic analysis of diversity, population structure, virulence, and antimicrobial resistance in *Klebsiella pneumoniae*, an urgent threat to public health. *Proc Natl Acad Sci USA* **112**: E3574-E3581.

House of Commons Committee. (2024). *The Antimicrobial Potential of Bacteriophages*. Parliament U: London.

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

Hsieh PF, Lin TL, Yang FL, Wu MC, Pan YJ, Wu SH and Wang JT. (2012). Lipopolysaccharide O1 antigen contributes to the virulence in *Klebsiella pneumoniae* causing pyogenic liver abscess. *PLoS One* **7**: e33155.

Hsu BB, Gibson TE, Yeliseyev V, Liu Q, Lyon L, Bry L, Silver PA and Gerber GK. (2019). Dynamic modulation of the gut microbiota and metabolome by bacteriophages in a mouse model. *Cell Host Microbe* **25**: 803-814.

Hu B, Margolin W, Molineux IJ and Liu J. (2015). Structural remodeling of bacteriophage T4 and host membranes during infection initiation. *Proc Natl Acad Sci USA* **112**: E4919-E4928.

Hu Y, Wei L, Feng Y, Xie Y and Zong Z. (2019). *Klebsiella huaxiensis* sp. nov., recovered from human urine. *Int J Syst Evol Microbiol* **69**: 333-336.

Hubert B. (2022). SkewDB, a comprehensive database of GC and 10 other skews for over 30,000 chromosomes and plasmids. *Sci Data* **9**: 92.

Hutchings MI, Truman AW and Wilkinson B. (2019). Antibiotics: past, present and future. *Curr Opin Microbiol* **51**: 72-80.

ICTV. (2023). Current ICTV Taxonomy Release. Available from: https://ictv.global/taxonomy. (Accessed: 04 Aug 2023).

ICTV Executive Committee. (2020). The new scope of virus taxonomy: partitioning the virosphere into 15 hierarchical ranks. *Nat Microbiol* **5**: 668-674.

Izdebski R, Fiett J, Urbanowicz P*, et al.* (2015). Phylogenetic lineages, clones and β-lactamases in an international collection of *Klebsiella oxytoca* isolates non-susceptible to expanded-spectrum cephalosporins. *J Antimicrob Chemother* **70**: 3230-3237.

Jacoby GA. (2009). AmpC β-lactamases. *Clin Microbiol Rev* **22**: 161-182, Table of Contents.

Jaeger T and Mayer C. (2008). The transcriptional factors MurR and catabolite activator protein regulate N-acetylmuramic acid catabolism in *Escherichia coli*. *J Bacteriol* **190**: 6598-6608.

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

James CE, Mahendran KR, Molitor A, Bolla JM, Bessonov AN, Winterhalter M and Pagès JM. (2009). How beta-lactam antibiotics enter bacteria: a dialogue with the porins. *PLoS One* **4**: e5453.

Jana B, Cain AK, Doerrler WT, Boinett CJ, Fookes MC, Parkhill J and Guardabassi L. (2017). The secondary resistome of multidrug-resistant Klebsiella pneumoniae. *Sci Rep* **7**: 42483.

Jansen M, Wahida A, Latz S, Krüttgen A, Häfner H, Buhl EM, Ritter K and Horz H-P. (2018). Enhanced antibacterial effect of the novel T4-like bacteriophage KARL-1 in combination with antibiotics against multi-drug resistant *Acinetobacter baumannii*. *Sci Rep* **8**: 14140.

Janssen AB, Doorduijn DJ, Mills G, Rogers MRC, Bonten MJM, Rooijakkers SHM, Willems RJL, Bengoechea JA and van Schaik W. (2020). Evolution of colistin resistance in the *Klebsiella pneumoniae* complex follows multiple evolutionary trajectories with variable effects on fitness and virulence characteristics. *Antimicrob Agents Chemother* **65**: e01958-01920.

Jault P, Leclerc T, Jennes S*, et al.* (2019). Efficacy and tolerability of a cocktail of bacteriophages to treat burn wounds infected by *Pseudomonas aeruginosa* (PhagoBurn): a randomised, controlled, double-blind phase 1/2 trial. *Lancet Infect Dis* **19**: 35-45.

Jindal S, Yang L, Day PJ and Kell DB. (2019). Involvement of multiple influx and efflux transporters in the accumulation of cationic fluorescent dyes by *Escherichia coli. BMC Microbiol* **19**: 195.

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

Jolley KA, Bliss CM, Bennett JS*, et al.* (2012). Ribosomal multilocus sequence typing: universal characterization of bacteria from domain to strain. *Microbiology* **158**: 1005-1015.

Jones JD, Trippett C, Suleman M, Clokie MRJ and Clark JR. (2023). The future of clinical phage therapy in the United Kingdom. Vol. 15 pp.

Joudeh N, Saragliadis A, Schulz C, Voigt A, Almaas E and Linke D. (2021). Transcriptomic response analysis of *Escherichia coli* to palladium stress. *Front Microbiol* **12**: 741836.

Kalischuk M, Hachey J, Thomas D, Johnson D and Kawchuk L. (2019). Complete genome sequence of phytopathogenic *Pectobacterium atrosepticum* lytic bacteriophage Peat2. *Am J Potato Res* **96**: 614-616.

Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A and Jermiin LS. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods* **14**: 587-589.

Katoh K and Standley DM. (2013). MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol Biol Evol* **30**: 772-780.

Kelly L and Jameson E. (2023). Bacteriophage cocktail shows no toxicity and improves the survival of *Galleria mellonella* infected with *Klebsiella* spp. *bioRxiv*.

Kęsik-Szeloch A, Drulis-Kawa Z, Weber-Dąbrowska B, Kassner J, Majkowska-Skrobek G, Augustyniak D, Łusiak-Szelachowska M, Żaczek M, Górski A and Kropinski AM. (2013). Characterising the biology of novel lytic bacteriophages infecting multidrug resistant *Klebsiella pneumoniae. Virol J* **10**: 100.

Khan MSI, Gao X, Liang K, Mei S and Zhan J. (2021). Virulent *Drexlervirial* bacteriophage MSK, morphological and genome resemblance with Rtp bacteriophage inhibits the multidrug-resistant bacteria. *Front Microbiol* **12**: 706700.

Kienesberger S, Cosic A, Kitsera M*, et al.* (2022). Enterotoxin tilimycin from gut-resident *Klebsiella* promotes mutational evolution and antibiotic resistance in mice. *Nat Microbiol* **7**: 1834-1848.

Kim Y, Lee SM, Nong LK, Kim J, Kim SB and Kim D. (2022). Characterization of *Klebsiella pneumoniae* bacteriophages, KP1 and KP12, with deep learning-based structure prediction. *Front Microbiol* **13**: 990910.

Koberg S, Brinks E, Albrecht V, Neve H and Franz C. (2018). Complete genome sequence of the novel virulent phage PMBT28 with lytic activity against thermotolerant *Salmonella enterica* subsp. *enterica* serovar Senftenberg ATCC 43845. *Genome Announc* **6**: e00568-00518.

Koberg S, Brinks E, Fiedler G, Hüsing C, Cho G-S, Hoeppner Marc P, Heller Knut J, Neve H and Franz Charles MAP. (2017). Genome sequence of *Klebsiella pneumoniae* bacteriophage PMBT1 isolated from raw sewage. *Genome Announc* **5**: e00914-00916.

Kongari R, Rajaure M, Cahill J, Rasche E, Mijalis E, Berry J and Young R. (2018). Phage spanins: diversity, topological dynamics and gene convergence. *BMC Bioinform* **19**: 326.

Konstantinidis KT and Tiedje JM. (2005). Genomic insights that advance the species definition for prokaryotes. *Proc Natl Acad Sci USA* **102**: 2567-2572.

Koonin EV, Dolja VV, Krupovic M, Varsani A, Wolf YI, Yutin N, Zerbini FM and Kuhn JH. (2020). Global organization and proposed megataxonomy of the vrus world. *Microbiol Mol Biol Rev* **84**.

Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH and Phillippy AM. (2017). Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res* **27**: 722-736.

Kortright KE, Chan BK and Turner PE. (2020). High-throughput discovery of phage receptors using transposon insertion sequencing of bacteria. *Proc Natl Acad Sci USA* **117**: 18670-18679.

Kortright KE, Chan BK, Koff JL and Turner PE. (2019). Phage therapy: a renewed approach to combat antibiotic-resistant bacteria. *Cell Host Microbe* **25**: 219-232.

Kortright KE, Doss-Gollin S, Chan BK and Turner PE. (2021). Evolution of bacterial cross-resistance to lytic phages and albicidin antibiotic. *Front Microbiol* **12**: 658374.

Kot W, Vogensen FK, Sørensen SJ and Hansen LH. (2014). DPS – A rapid method for genome sequencing of DNA-containing bacteriophages directly from a single plaque. *J Virol Methods* **196**: 152-156.

Kramberger P, Honour RC, Herman RE, Smrekar F and Peterka M. (2010). Purification of the *Staphylococcus aureus* bacteriophages VDX-10 on methacrylate monoliths. *J Virol Methods* **166**: 60-64.

Krause KM, Serio AW, Kane TR and Connolly LE. (2016). Aminoglycosides: An overview. *Cold Spring Harb Perspect Med* **6**: a027029.

Kropinski AM, Mazzocco A, Waddell TE, Lingohr E and Johnson RP. (2009). Enumeration of bacteriophages by double agar overlay plaque assay. *Methods Mol Biol* **501**: 69-76.

Ku H, Kabwe M, Chan HT, Stanton C, Petrovski S, Batinovic S and Tucci J. (2021). Novel *Drexlerviridae* bacteriophage KMI8 with specific lytic activity against *Klebsiella michiganensis* and its biofilms. *PLoS One* **16**: e0257102.

Kumar S, Stecher G, Li M, Knyaz C and Tamura K. (2018). MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol* **35**: 1547-1549.

Kutter E. (2005). Phage host range and efficiency of plating. *Bacteriophages: Biology and Applications,* (Kutter E & Sulakvelidze A, eds.), pp. 141-149. CRC Press, Boca Raton, Florida.

Kutter E, Raya R and Carlson K. (2005). Molecular mechanisms of phage infection. *Bacteriophages: Biology and Applications,* (Kutter E & Sulakvelidze A, eds.), pp. 165-222. CRC Press, Boca Raton, Florida.

Kwon HJ, Cho SH, Kim TE, Won YJ, Jeong J, Park SC, Kim JH, Yoo HS, Park YH and Kim SJ. (2008). Characterization of a T7-like lytic bacteriophage (phiSG-JL2) of *Salmonella enterica* serovar Gallinarum biovar Gallinarum. *Appl Environ Microbiol* **74**: 6970-6979.

Kwon T, Jung YH, Lee S, Yun MR, Kim W and Kim DW. (2016). Comparative genomic analysis of *Klebsiella pneumoniae* subsp. *pneumoniae* KP617 and PittNDM01, NUHL24835, and ATCC BAA-2146 reveals unique evolutionary history of this strain. *Gut Pathog* **8**: 34.

Labudda Ł, Strapagiel D, Karczewska-Golec J and Golec P. (2017). Complete annotated genome sequences of four *Klebsiella pneumoniae* phages isolated from sewage in Poland. *Genome Announc* **5**.

Lam MMC, Wick RR, Judd LM, Holt KE and Wyres KL. (2022). Kaptive 2.0: updated capsule and lipopolysaccharide locus typing for the *Klebsiella pneumoniae* species complex. *Microb Genom* **8**.

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

Lam MMC, Wyres KL, Wick RR, Judd LM, Fostervold A, Holt KE and Löhr IH. (2019). Convergence of virulence and MDR in a single plasmid vector in MDR *Klebsiella pneumoniae* ST15. *J Antimicrob Chemother* **74**: 1218-1222.

Langmead B and Salzberg SL. (2012). Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**: 357-359.

Langridge GC, Phan MD, Turner DJ*, et al.* (2009). Simultaneous assay of every *Salmonella* Typhi gene using one million transposon mutants. *Genome Res* **19**: 2308-2316.

Larsen F, Offersen SM, Li VR, Deng L, Nielsen DS and Rasmussen TS. (2023). Choice of ultrafilter affects recovery rate of bacteriophages. *Viruses* **15**.

Latka A, Leiman PG, Drulis-Kawa Z and Briers Y. (2019). Modeling the architecture of depolymerase-containing receptor binding proteins in *Klebsiella* Phages. *Front Microbiol* **10**: 2649.

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

Lavigne R, Seto D, Mahadevan P, Ackermann H-W and Kropinski AM. (2008). Unifying classical and molecular taxonomic classification: analysis of the *Podoviridae* using BLASTP-based tools. *Res Microbiol* **159**: 406-414.

Lawlor MS, O'Connor C and Miller VL. (2007). Yersiniabactin is a virulence factor for *Klebsiella pneumoniae* during pulmonary infection. *Infect Immun* **75**: 1463-1472.

Lawlor MS, Hsu J, Rick PD and Miller VL. (2005). Identification of *K. pneumoniae* virulence determinants using an intranasal infection model. *Mol Microbiol* **58**: 1054-1073.

Lee JH, Cheon IS, Shim BS, Kim DW, Kim SW, Chun J and Song M. (2012). Draft genome sequence of *Klebsiella pneumoniae* subsp. *pneumoniae* DSM 30104T. *J Bacteriol* **194**: 5722-5723.

Lee JY, Kong M, Oh J, Lim J, Chung SH, Kim J-M, Kim J-S, Kim K-H, Yoo J-C and Kwak W. (2021). Comparative evaluation of Nanopore polishing tools for microbial genome assembly and polishing strategies for downstream analysis. *Sci Rep* **11**: 20740.

Leggett RM, Alcon-Giner C, Heavens D*, et al.* (2020). Rapid MinION profiling of preterm microbiota and antimicrobial-resistant pathogens. *Nat Microbiol* **5**: 430-442.

Leiman PG, Chipman PR, Kostyuchenko VA, Mesyanzhinov VV and Rossmann MG. (2004). Three-dimensional rearrangement of proteins in the tail of bacteriophage T4 on infection of its host. *Cell* **118**: 419-429.

Lenski RE, Wiser MJ, Ribeck N*, et al.* (2015). Sustained fitness gains and variability in fitness trajectories in the long-term evolution experiment with *Escherichia coli*. *Proc Biol Sci* **282**: 20152292.

Lepper PM, Held TK, Schneider EM, Bölke E, Gerlach H and Trautmann M. (2002). Clinical implications of antibiotic-induced endotoxin release in septic shock. *Intensive Care Med* **28**: 824-833.

Letunic I and Bork P. (2021). Interactive tree of life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res*.

Li B, Zhao Y, Chen Z and Zhou D. (2014). Molecular pathogenesis of *Klebsiella pneumoniae*. *Future Microbiol* **9**: 1071-1081.

Li H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv*.

Li J, Zhang H, Ning J, Sajid A, Cheng G, Yuan Z and Hao H. (2019). The nature and epidemiology of OqxAB, a multidrug efflux pump. *Antimicrob Resist Infect Control* **8**: 44-57.

Li S, Fan H, An X, Fan H, Jiang H, Chen Y and Tong Y. (2014). Scrutinizing virus genome termini by high-throughput sequencing. *PLoS One* **9**: e85806.

Lin L, Wei C, Chen M, Wang H, Li Y, Li Y, Yang L and An Q. (2015). Complete genome sequence of endophytic nitrogen-fixing *Klebsiella variicola* strain DX120E. *Stand in Genomic Sci* **10**: 22.

Lin RCY, Sacher JC, Ceyssens P-J, Zheng J, Khalid A and Iredell JR. (2021). Phage biobank: present challenges and future perspectives. *Curr Opin Biotechnol* **68**: 221-230.

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

Lin TL, Yang FL, Ren CT*, et al.* (2022). Development of *Klebsiella pneumoniae* capsule polysaccharide-conjugated vaccine candidates using phage depolymerases. *Front Immunol* **13**: 843183.

Lin Y, Yuan J, Kolmogorov M, Shen MW, Chaisson M and Pevzner PA. (2016). Assembly of long error-prone reads using de Bruijn graphs. *Proc Natl Acad Sci USA* **113**: E8396-E8405.

Lin Y, Chang RYK, Britton WJ, Morales S, Kutter E and Chan HK. (2018). Synergy of nebulized phage PEV20 and ciprofloxacin combination against *Pseudomonas aeruginosa*. *Int J Pharm* **551**: 158-165.

Little JW. (2006). Gene regulatory circuitry of phage λ. *The Bacteriophages,* (Calendar R, ed.) pp. 74-82. Oxford University Press, Oxford.

Liu B, Zheng D, Jin Q, Chen L and Yang J. (2018). VFDB 2019: a comparative pathogenomic platform with an interactive web interface. *Nucleic Acids Res* **47**: D687-D692.

Liu GF, Wang XX, Su HZ and Lu GT. (2021). Progress on the GntR family transcription regulators in bacteria. *Yi Chuan* **43**: 66-73.

Liu R and Ochman H. (2007). Origins of flagellar gene operons and secondary flagellar systems. *J Bacteriol* **189**: 7098-7104.

Liu Y, Leung SSY, Huang Y*, et al.* (2020). Identification of two depolymerases from phage IME205 and their antivirulent functions on K47 capsule of *Klebsiella pneumoniae*. *Front Microbiol* **11**: 218.

Long SW, Linson SE, Ojeda Saavedra M, Cantu C, Davis JJ, Brettin T and Olsen RJ. (2017). Whole-genome sequencing of human clinical *Klebsiella pneumoniae* isolates reveals misidentification and misunderstandings of *Klebsiella pneumoniae*, *Klebsiella variicola*, and *Klebsiella quasipneumoniae*. *mSphere* **2**: e00290-00217.

Lourenço M, Osbelt L, Passet V, Gravey F, Megrian D, Strowig T, Rodrigues C and Brisse S. (2023). Phages against noncapsulated *Klebsiella pneumoniae*: broader host range, slower resistance. *Microbiol Spectr* **0**: e04812-04822.

Low WW, Wong JLC, Beltran LC*, et al.* (2022). Mating pair stabilization mediates bacterial conjugation species specificity. *Nat Microbiol* **7**: 1016-1027.

Lu IN, Muller CP and He FQ. (2020). Applying next-generation sequencing to unravel the mutational landscape in viral quasispecies. *Virus Res* **283**: 197963.

Luong T, Salabarria A-C, Edwards RA and Roach DR. (2020). Standardized bacteriophage purification for personalized phage therapy. *Nat Protoc*.

Łusiak-Szelachowska M, Żaczek M, Weber-Dąbrowska B*, et al.* (2014). Phage neutralization by sera of patients receiving phage therapy. *Viral Immunol* **27**: 295-304.

Lv L, Wan M, Wang C*, et al.* (2020). Emergence of a plasmid-encoded resistance-nodulation-division efflux pump conferring resistance to multiple drugs, including tigecycline, in *Klebsiella pneumoniae*. *mBio* **11**: e02930-02919.

Ma Y, Wu X, Li S, Tang L, Chen M and An Q. (2021). Proposal for reunification of the genus *Raoultella* with the genus *Klebsiella* and reclassification of *Raoultella electrica* as *Klebsiella electrica* comb. nov. *Res Microbiol* **172**: 103851.

Maatallah M, Vading M, Kabir MH, Bakhrouf A, Kalin M, Naucler P, Brisse S and Giske CG. (2014). *Klebsiella variicola* is a frequent cause of bloodstream infection in the Stockholm area, and associated with higher mortality compared to *K. pneumoniae*. *PLoS One* **9**: e113539.

Maffei E, Shaidullina A, Burkolter M*, et al.* (2021). Systematic exploration of *Escherichia coli* phage-host interactions with the BASEL phage collection. *PLoS Biol* **19**: e3001424.

Maffei E, Woischnig A-K, Burkolter MR*, et al.* (2024). Phage Paride can kill dormant, antibiotic-tolerant cells of *Pseudomonas aeruginosa* by direct lytic replication. *Nat Commun* **15**: 175.

Magiorakos AP, Srinivasan A, Carey RB, *et al.* (2012). Multidrug-resistant, extensively drug-resistant and pandrug-resistant bacteria: an international expert proposal for interim standard definitions for acquired resistance. *Clin Microbiol Infect* **18**: 268-281.

Maier L, Goemans CV, Wirbel J, *et al.* (2021). Unravelling the collateral damage of antibiotics on gut bacteria. *Nature* **599**: 120-124.

Majewska J, Beta W, Lecion D, *et al.* (2015). Oral application of T4 phage induces weak antibody production in the gut and in the blood. *Viruses* **7**: 4783-4799.

Malek A, McGlynn K, Taffner S, *et al.* (2019). Next-generation-sequencing-based hospital outbreak investigation yields insight into *Klebsiella aerogenes* population structure and determinants of carbapenem resistance and pathogenicity. *Antimicrob Agents Chemother* **63**: e02577-02518.

Mallawaarachchi V, Roach MJ, Decewicz P, *et al.* (2023). Phables: from fragmented assemblies to high-quality bacteriophage genomes. *Bioinformatics* **39**: btad586.

Malmir S, Bahreinian M, Zahiri Yeganeh S, Mirnejad R, Moosazadeh Moghaddam M and Saberi F. (2018). Molecular mechanisms of resistance to conventional antibiotics in bacteria. *Int J Med Rev* **5**: 118-129.

Manandhar S, Nguyen Q, Nguyen Thi Nguyen T, Pham DT, Rabaa MA, Dongol S, Basnyat B, Dixit SM, Baker S and Karkey A. (2022). Genomic epidemiology, antimicrobial resistance and virulence factors of *Enterobacter cloacae* complex causing potential community-onset bloodstream infections in a tertiary care hospital of Nepal. *JAC Antimicrob Resist* **4**.

Mangalea MR and Duerkop BA. (2020). Fitness trade-offs resulting from bacteriophage resistance potentiate synergistic antibacterial strategies. *Infect Immun* **88**.

Mäntynen S, Laanto E, Oksanen HM, Poranen MM and Díaz-Muñoz SL. (2021). Black box of phage-bacterium interactions: exploring alternative phage infection strategies. *Open Biol* **11**: 210188.

Martin RM and Bachman MA. (2018). Colonization, infection, and the accessory genome of *Klebsiella pneumoniae*. *Front Cell Infect Microbiol* **8**.

Matilla MA, Fang X and Salmond GP. (2014). Viunalikeviruses are environmentally common agents of horizontal gene transfer in pathogens and biocontrol bacteria. *ISME J* **8**: 2143-2147.

Maurya AP, Dhar D, Basumatary MK, Paul D, Ingti B, Choudhury D, Talukdar AD, Chakravarty A, Mishra S and Bhattacharjee A. (2017). Expansion of highly stable bla_{OXA-10} β-lactamase family within diverse host range among nosocomial isolates of Gram-negative bacilli within a tertiary referral hospital of Northeast India. *BMC Res Notes* **10**: 145-150.

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

McDougall FK, Wyres KL, Judd LM, Boardman WSJ, Holt KE and Power ML. (2021). Novel strains of *Klebsiella africana* and *Klebsiella pneumoniae* in Australian fruit bats (*Pteropus poliocephalus*). *Res Microbiol* **172**: 103879.

McLennan DA. (2010). How to read a phylogenetic tree. *Evolution: Education and Outreach* **3**: 506-519.

Meile S, Du J, Staubli S, *et al.* (2023). Engineered reporter phages for detection of *Escherichia coli*, *Enterococcus*, and *Klebsiella* in urine. *Nat Commun* **14**: 4336.

Mentasti M, David S, Turton J, *et al.* (2023). Clonal expansion and rapid characterization of *Klebsiella pneumoniae* ST1788, an otherwise uncommon strain spreading in Wales, UK. *Microb Genom* **9**: 001104.

Merabishvili M, Pirnay J-P and De Vos D. (2018). Guidelines to compose an ideal bacteriophage cocktail. *Bacteriophage Therapy: From Lab to Clinical Practice,* (Azeredo J & Sillankorva S, eds.), pp. 99-110. Springer New York, New York, NY.

Merla C, Rodrigues C, Passet V, *et al.* (2019). Description of *Klebsiella spallanzanii* sp. nov. and of *Klebsiella pasteurii* sp. nov. *Front Microbiol* **10**: 2360.

Merrill BD, Ward AT, Grose JH and Hope S. (2016). Software-based analysis of bacteriophage genomes, physical ends, and packaging strategies. *BMC Genomics* **17**: 679.

Mijalis EM, Lessor LE, Cahill JL, Rasche ES and Kuty Everett GF. (2015). Complete genome sequence of *Klebsiella pneumoniae* carbapenemase-producing *K. pneumoniae* myophage Miro. *Genome Announc* **3**: e01137-01115.

Miller ES, Kutter E, Mosig G, Arisaka F, Kunisawa T and Rüger W. (2003). Bacteriophage T4 genome. *Microbiol Mol Biol Rev* **67**: 86-156.

Minato Y, Dawadi S, Kordus SL, Sivanandam A, Aldrich CC and Baughn AD. (2018). Mutual potentiation drives synergy between trimethoprim and sulfamethoxazole. *Nat Commun* **9**: 1003.

Molina F, Menor-Flores M, Fernández L, Vega-Rodríguez MA and García P. (2022). Systematic analysis of putative phage-phage interactions on minimum-sized phage cocktails. *Sci Rep* **12**: 2458.

Molineux IJ. (2006). Fifty-three years since Hershey and Chase; much ado about pressure but which pressure is it? *Virology* **344**: 221-229.

Molineux IJ. (2006). The T7 group. *The Bacteriophages,* (Calendar R, ed.) pp. 277-301. Oxford University Press, Oxford.

Moore SD and Prevelige PE. (2002). DNA packaging: a new class of molecular motors. *Curr Biol* **12**: R96-R98.

Moradigaravand D, Martin V, Peacock SJ and Parkhill J. (2017). Evolution and epidemiology of multidrug-resistant *Klebsiella pneumoniae* in the United Kingdom and Ireland. *mBio* **8**: e01976-01916.

Moraru C, Varsani A and Kropinski AM. (2020). VIRIDIC - a novel tool to calculate the intergenomic similarities of prokaryote-infecting viruses. *Viruses* **12**: 1268.

Moreno-Vivián C, Cabello P, Martínez-Luque M, Blasco R and Castillo F. (1999). Prokaryotic nitrate reduction: molecular properties and functional distinction among bacterial nitrate reductases. *J Bacteriol* **181**: 6573-6584.

Morgan AD, Bonsall MB and Buckling A. (2010). Impact of bacterial mutation rate on coevolutionary dynamics between bacteria and phages. *Evolution* **64**: 2980-2987.

Mushegian AR. (2020). Are there 10^{31} virus particles on Earth, or more, or fewer? *J Bacteriol* **202**: e00052-00020.

Nagel T, Musila L, Muthoni M, Nikolich M, Nakavuma JL and Clokie MR. (2022). Phage banks as potential tools to rapidly and cost-effectively manage antimicrobial resistance in the developing world. *Curr Opin Virol* **53**: 101208.

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

Nayfach S, Camargo AP, Schulz F, Eloe-Fadrosh E, Roux S and Kyrpides NC. (2021). CheckV assesses the quality and completeness of metagenome-assembled viral genomes. *Nat Biotechnol* **39**: 578-585.

Nguyen L-T, Schmidt HA, von Haeseler A and Minh BQ. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* **32**: 268-274.

NICE. (2014). *Infection Prevention and Control: NICE Quality Standard (QS61)*. NICE: London.

Nieweg A and Bremer E. (1997). The nucleoside-specific Tsx channel from the outer membrane of *Salmonella* typhimurium, *Klebsiella pneumoniae* and *Enterobacter aerogenes*: functional characterization and DNA sequence analysis of the tsx genes. *Microbiology (Reading)* **143 (Pt 2)**: 603-615.

Nikolic I, Vukovic D, Gavric D, Cvetanovic J, Aleksic Sabo V, Gostimirovic S, Narancic J and Knezevic P. (2022). An optimized checkerboard method for phage-antibiotic synergy detection. *Viruses* **14**.

Nilsson AS and Ljungquist EH. (2006). The P2-like bacteriophages. *The Bacteriophages,* (Calendar R, ed.) pp. 365-390. Oxford University Press, Oxford.

Nishimura Y, Yoshida T, Kuronishi M, Uehara H, Ogata H and Goto S. (2017). ViPTree: the viral proteomic tree server. *Bioinformatics* **33**: 2379-2380.

Niu H, Li T, Du Y, Lv Z, Cao Q and Zhang Y. (2023). Glutamate transporters GltS, GltP and GltI are involved in *Escherichia coli* tolerance in vitro and pathogenicity in mouse urinary tract infections. *Microorganisms* **11**: 1173.

Nobrega FL, Vlot M, de Jonge PA, Dreesens LL, Beaumont HJE, Lavigne R, Dutilh BE and Brouns SJJ. (2018). Targeting mechanisms of tailed bacteriophages. *Nat Rev Microbiol* **16**: 760-773.

Nordstrom HR, Evans DR, Finney AG, *et al.* (2022). Genomic characterization of lytic bacteriophages targeting genetically diverse *Pseudomonas aeruginosa* clinical isolates. *iScience* **25**: 104372.

Ofir G and Sorek R. (2018). Contemporary phage biology: from classic models to new insights. *Cell* **172**: 1260-1270.

Oh HK, Jo JH, Hwang YJ and Myung H. (2020). Complete genome sequence of a novel bacteriophage, ATCEA85, infecting *Enterobacter aerogenes. Arch Virol* **165**: 2397-2400.

Okonechnikov K, Golosova O, Fursov M and team U. (2012). Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics* **28**: 1166-1167.

Owen SV, Perez-Sepulveda BM and Adriaenssens EM. (2021). Detection of bacteriophages: sequence-based systems. *Bacteriophages: Biology, Technology, Therapy,* (Harper D, Abedon S, Burrowes B & McConville M, eds.), pp. 621–644. Springer Nature Switzerland AG, New York City, NY.

Paczosa MK and Mecsas J. (2016). *Klebsiella pneumoniae*: going on the offense with a strong defense. *Microbiol Mol Biol Rev* **80**: 629-661.

Paczosa MK, Silver RJ, McCabe AL, Tai AK, McLeish CH, Lazinski DW and Mecsas J. (2020). Transposon mutagenesis screen of *Klebsiella pneumoniae* identifies multiple genes important for resisting antimicrobial activities of neutrophils in mice. *Infect Immun* **88**: e00034-00020.

Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MTG, Fookes M, Falush D, Keane JA and Parkhill J. (2015). Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics* **31**: 3691-3693.

Pagès JM, James CE and Winterhalter M. (2008). The porin and the permeating antibiotic: a selective diffusion barrier in Gram-negative bacteria. *Nat Rev Microbiol* **6**: 893-903.

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

Pan YJ, Lin TL, Chen CC, Tsai YT, Cheng YH, Chen YY, Hsieh PF, Lin YT and Wang JT. (2017). *Klebsiella* phage PhiK64-1 encodes multiple depolymerases for multiple host capsular types. *J Virol* **91**: e02457-02416.

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

Pan YJ, Lin TL, Lin YT, Su PA, Chen CT, Hsieh PF, Hsu CR, Chen CC, Hsieh YC and Wang JT. (2015). Identification of capsular types in carbapenem-resistant *Klebsiella pneumoniae* strains by *wzc* sequencing and implications for capsule depolymerase treatment. *Antimicrob Agents Chemother* **59**: 1038-1047.

Park SJ and Gunsalus RP. (1995). Oxygen, iron, carbon, and superoxide control of the fumarase fumA and fumC genes of *Escherichia coli*: role of the arcA, fnr, and soxR gene products. *J Bacteriol* **177**: 6255-6262.

Partridge JD, Bodenmiller DM, Humphrys MS and Spiro S. (2009). NsrR targets in the *Escherichia coli* genome: new insights into DNA sequence requirements for binding and a role for NsrR in the regulation of motility. *Mol Microbiol* **73**: 680-694.

Passet V and Brisse S. (2018). Description of *Klebsiella grimontii* sp. nov. *Int J Syst Evol Microbiol* **68**: 377-381.

Patro LPP and Rathinavelan T. (2019). Targeting the sugary armor of *Klebsiella* species. *Front Cell Infect Microbiol* **9**: 367.

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

Pearce ME, Langridge GC, Lauer AC, Grant K, Maiden MCJ and Chattaway MA. (2021). An evaluation of the species and subspecies of the genus *Salmonella* with whole genome sequence data: proposal of type strains and epithets for novel *S. enterica* subspecies VII, VIII, IX, X and XI. *Genomics* **113**: 3152-3162.

Pelfrene E, Willebrand E, Cavaleiro Sanches A, Sebris Z and Cavaleri M. (2016). Bacteriophage therapy: a regulatory perspective. *J Antimicrob Chemother* **71**: 2071-2074.

Peng Q, Fang M, Liu X, Zhang C, Liu Y and Yuan Y. (2020). Isolation and characterization of a novel phage for controlling multidrug-resistant *Klebsiella pneumoniae*. *Microorganisms* **8**: 542.

Peng Q, Ma Z, Han Q, Xiang F, Wang L, Zhang Y, Zhao Y, Li J, Xian Y and Yuan Y. (2023). Characterization of bacteriophage vB_KleM_KB2 possessing high control ability to pathogenic *Klebsiella pneumoniae*. *Sci Rep* **13**: 9815.

Petit RA, 3rd and Read TD. (2020). Bactopia: a flexible pipeline for complete analysis of bacterial genomes. *mSystems* **5**.

Pieńko T and Trylska J. (2020). Extracellular loops of BtuB facilitate transport of vitamin B12 through the outer membrane of *E. coli*. *PLoS Comput Biol* **16**: e1008024.

Pires DP, Melo LDR and Azeredo J. (2021). Understanding the complex phage-host interactions in biofilm communities. *Annu Rev Virol* **8**: 73-94.

Pirnay J-P, Djebara S, Steurs G, *et al.* (2023). Retrospective, observational analysis of the first one hundred consecutive cases of personalized bacteriophage therapy of difficult-to-treat infections facilitated by a Belgian consortium. *medRxiv*.

Pirnay JP, Verbeken G, Ceyssens PJ, Huys I, De Vos D, Ameloot C and Fauconnier A. (2018). The magistral phage. *Viruses* **10**: E64.

Pitout JD, Nordmann P and Poirel L. (2015). Carbapenemase-producing *Klebsiella pneumoniae*, a key pathogen set for global nosocomial dominance. *Antimicrob Agents Chemother* **59**: 5873-5884.

Podschun R and Ullman U. (1998). *Klebsiella* spp. as nosocomial pathogens: epidemiology, taxonomy, typing methods, and pathogenicity factors. *Clin Microbiol Rev* **11**: 589-603.

Pöltl L, Kitsera M, Raffl S, *et al.* (2023). Microbiota-derived genotoxin tilimycin generates colonic stem cell mutations. *Cell Rep* **42**: 112199.

Pons BJ, Dimitriu T, Westra ER and van Houte S. (2023). Antibiotics that affect translation can antagonize phage infectivity by interfering with the deployment of counter-defenses. *Proc Natl Acad Sci USA* **120**: e2216084120.

Povolotsky TL and Hengge R. (2015). Genome-based comparison of cyclic di-GMP signaling in pathogenic and commensal *Escherichia coli* strains. *J Bacteriol* **198**: 111-126.

Putze J, Hennequin C, Nougayrède J-P, *et al.* (2009). Genetic structure and distribution of the colibactin genomic island among members of the family *Enterobacteriaceae*. *Infect Immun* **77**: 4696-4703.

Queenan AM and Bush K. (2007). Carbapenemases: the versatile β-lactamases. *Clin Microbiol Rev* **20**: 440-458.

Rai AK and Mitchell AM. (2020). Enterobacterial common antigen: synthesis and function of an enigmatic molecule. *mBio* **11**: e01914-01920.

Rakov C, Ben Porat S, Alkalay-Oren S, *et al.* (2021). Targeting biofilm of MDR *Providencia stuartii* by phages using a catheter model. *Antibiotics (Basel)* **10**.

Ramirez MS and Tolmasky ME. (2010). Aminoglycoside modifying enzymes. *Drug Resist Updat* **13**: 151-171.

Rebula L, Raspor A, Bavčar M, Štrancar A and Leskovec M. (2023). CIM monolithic chromatography as a useful tool for endotoxin reduction and purification of bacteriophage particles supported with PAT analytics. *J Chromatogr B Analyt Technol Biomed Life Sci* **1217**: 123606.

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

Reznikoff WS. (2008). Transposon Tn5. *Annu Rev Genet* **42**: 269-286.

Roberts GA, Stephanou AS, Kanwar N, Dawson A, Cooper LP, Chen K, Nutley M, Cooper A, Blakely GW and Dryden DT. (2012). Exploring the DNA mimicry of the Ocr protein of phage T7. *Nucleic Acids Res* **40**: 8129-8143.

Rocha EP. (2004). Codon usage bias from tRNA's point of view: redundancy, specialization, and efficient decoding for translation optimization. *Genome Res* **14**: 2279-2286.

Rocha J, Henriques I, Gomila M and Manaia CM. (2022). Common and distinctive genomic features of *Klebsiella pneumoniae* thriving in the natural environment or in clinical settings. *Scientific Reports* **12**: 10441.

Rodionov DA. (2007). Comparative genomic reconstruction of transcriptional regulatory networks in bacteria. *Chem Rev* **107**: 3467-3497.

Rodrigues C, Lanza VF, Peixe L, Coque TM and Novais Â. (2023). Phylogenomics of globally spread clonal groups 14 and 15 of *Klebsiella pneumoniae*. *Microbiol Spectr* **11**: e0339522.

Rodrigues C, Passet V, Rakotondrasoa A, Diallo TA, Criscuolo A and Brisse S. (2019). Description of *Klebsiella africanensis* sp. nov., *Klebsiella variicola* subsp. *tropicalensis* subsp. nov. and *Klebsiella variicola* subsp. *variicola* subsp. nov. *Res Microbiol* **170**: 165-170.

Rohwer F and Edwards R. (2002). The phage proteomic tree: a genome-based taxonomy for phage. *J Bacteriol* **184**: 4529-4535.

Romero-Calle D, Guimarães Benevides R, Góes-Neto A and Billington C. (2019). Bacteriophages as alternatives to antibiotics in clinical care. *Antibiotics (Basel)* **8**: E138.

Roof WD, Horne SM, Young KD and Young R. (1994). *slyD*, a host gene required for phi X174 lysis, is related to the FK506-binding protein family of peptidyl-prolyl cis-trans-isomerases. *J Biol Chem* **269**: 2902-2910.

Rosenblueth M, Martínez L, Silva J and Martínez-Romero E. (2004). *Klebsiella variicola*, a novel species with clinical and plant-associated isolates. *Syst Appl Microbiol* **27**: 27-35.

Roshankhah R, Jackson K, Nguyen TTN, Pelton R, Hosseinidoust Z and Ghosh R. (2023). Purification of phage for therapeutic applications using high throughput anion exchange membrane chromatography. *J Chromatogr B* **1229**: 123867.

Ross A, Ward S and Hyman P. (2016). More is better: selecting for broad host range bacteriophages. *Front Microbiol* **7**: 1352.

Roulston KJ, Bharucha T, Turton JF, Hopkins KL and Mack DJF. (2018). A case of NDM-carbapenemase-producing hypervirulent *Klebsiella pneumoniae* sequence type 23 from the UK. *JMM Case Rep* **5**: e005130.

Rozwandowicz M, Brouwer MSM, Fischer J, Wagenaar JA, Gonzalez-Zorn B, Guerra B, Mevius DJ and Hordijk J. (2018). Plasmids carrying antimicrobial resistance genes in Enterobacteriaceae. *J Antimicrob Chemother* **73**: 1121-1137.

Rudolph C, Freund-Mölbert E and Stirm S. (1975). Fragments of *Klebsiella* bacteriophage no. 11. *Virology* **64**: 236-246.

Sabnis A, Hagart KL, Klöckner A, *et al.* (2021). Colistin kills bacteria by targeting lipopolysaccharide in the cytoplasmic membrane. *Elife* **10**.

Sacher J and Zheng J. (2021). Phage therapy collaboration and compassionate use. *Bacteriophages: Biology, Technology, Therapy,* (Harper D, Abedon S, Burrowes B & McConville M, eds.), pp. 1069-1098. Springer Nature Switzerland AG, New York City, NY.

Sacher JC. (2023). Getting our AKTA-gether: using the AKTA Flux to purify phages for patients. Available from: https://phage.directory/capsid/akta-phage. (Accessed: 31 Aug 2023).

Sacher JC, Zheng J and Lin RCY. (2022). Data to power precision phage therapy: a look at the Phage Directory-Phage Australia Partnership. *Phage (New Rochelle)* **3**: 112-115.

Saha R, Farrance CE, Verghese B, Hong S and Donofrio RS. (2013). *Klebsiella michiganensis* sp. nov., a new bacterium isolated from a tooth brush holder. *Curr Microbiol* **66**: 72-78.

Savalia D, Robins W, Nechaev S, Molineux I and Severinov K. (2010). The role of the T7 Gp2 inhibitor of host RNA polymerase in phage development. *J Mol Biol* **402**: 118-126.

Sawa T, Kooguchi K and Moriyama K. (2020). Molecular diversity of extended-spectrum β-lactamases and carbapenemases, and antimicrobial resistance. *J Intensive Care* **8**: 13-25.

Scheu PD, Witan J, Rauschmeier M, Graf S, Liao Y-F, Ebert-Jung A, Basché T, Erker W and Unden G. (2012). CitA/CitB Two-component system regulating citrate fermentation in *Escherichia coli* and its relation to the DcuS/DcuR system *in vivo*. *J Bacteriol* **194**: 636-645.

Schmerer M, Molineux IJ and Bull JJ. (2014). Synergy as a rationale for phage therapy using phage cocktails. *PeerJ* **2**: e590.

Schwengers O, Hoek A, Fritzenwanker M, Falgenhauer L, Hain T, Chakraborty T and Goesmann A. (2020). ASA³P: An automatic and scalable pipeline for the assembly,

annotation and higher-level analysis of closely related bacterial isolates. *PLoS Comput Biol* **16**: e1007134.

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

Seemann T. (2016). ABRicate. Available from: https://github.com/tseemann/abricate. (Accessed: 18 Apr 2023).

Seemann T. (2018). Shovill. Available from: https://github.com/tseemann/shovill. (Accessed: 12 Jan 2021).

Seiffert SN, Wüthrich D, Gerth Y, Egli A, Kohler P and Nolte O. (2019). First clinical case of KPC-3-producing *Klebsiella michiganensis* in Europe. *New Microbes New Infect* **29**: 100516.

Seki M, Gotoh K, Nakamura S*, et al.* (2013). Fatal sepsis caused by an unusual *Klebsiella* species that was misidentified by an automated identification system. *J Med Microbiol* **62**: 801-803.

Sequeira RP, McDonald JAK, Marchesi JR and Clarke TB. (2020). Commensal Bacteroidetes protect against *Klebsiella pneumoniae* colonization and transmission through IL-36 signalling. *Nat Microbiol* **5**: 304-313.

Sereika M, Kirkegaard RH, Karst SM, Michaelsen TY, Sørensen EA, Wollenberg RD and Albertsen M. (2022). Oxford Nanopore R10.4 long-read sequencing enables the generation of near-finished bacterial genomes from pure cultures and metagenomes without short-read or reference polishing. *Nature Methods* **19**: 823-826.

Severinova E and Severinov K. (2006). Localization of the *Escherichia coli* RNA polymerase beta' subunit residue phosphorylated by bacteriophage T7 kinase Gp0.7. *J Bacteriol* **188**: 3470-3476.

Shankar C, Jacob JJ, Sugumar SG*, et al.* (2021). Distinctive mobile genetic elements observed in the clonal expansion of carbapenem-resistant *Klebsiella pneumoniae* in India. *Microb Drug Resist* **27**: 1096-1104.

Shin SH, Kim S, Kim JY, Lee S, Um Y, Oh MK, Kim YR, Lee J and Yang KS. (2012). Complete genome sequence of *Enterobacter aerogenes* KCTC 2190. *J Bacteriol* **194**: 2373-2374.

Short FL, Di Sario G, Reichmann NT, Kleanthous C, Parkhill J and Taylor PW. (2020). Genomic profiling reveals distinct routes to complement resistance in *Klebsiella pneumoniae*. *Infect Immun* **88**.

Sitto F and Battistuzzi FU. (2019). Estimating pangenomes with Roary. *Mol Biol Evol* **37**: 933-939.

Smith-Zaitlik T, Shibu P, McCartney AL, Foster G, Hoyles L and Negus D. (2022). Extended genomic analyses of the broad-host-range phages vB_KmiM-2Di and vB_KmiM-4Dii reveal slopekviruses have highly conserved genomes. *Microbiology* **168**.

Smrekar F, Ciringer M, Strancar A and Podgornik A. (2011). Characterisation of methacrylate monoliths for bacteriophage purification. *J Chromatogr A* **1218**: 2438-2444.

Smrekar F, Ciringer M, Peterka M, Podgornik A and Strancar A. (2008). Purification and concentration of bacteriophage T4 using monolithic chromatographic supports. *J Chromatogr B Analyt Technol Biomed Life Sci* **861**: 177-180.

Sodergren EJ, Hsu PY and DeMoss JA. (1988). Roles of the *narJ* and *narI* gene products in the expression of nitrate reductase in *Escherichia coli*. *J Biol Chem* **263**: 16156-16162.

Ssekatawa K, Byarugaba DK, Nakavuma JL, Kato CD, Ejobi F, Tweyongyere R and Eddie WM. (2021). Prevalence of pathogenic *Klebsiella pneumoniae* based on PCR capsular typing harbouring carbapenemases encoding genes in Uganda tertiary hospitals. *Antimicrob Resist Infect Control* **10**: 57.

Stahlhut SG, Struve C, Krogfelt KA and Reisner A. (2012). Biofilm formation of *Klebsiella pneumoniae* on urethral catheters requires either type 1 or type 3 fimbriae. *FEMS Immunol Med Microbiol* **65**: 350-359.

Stoesser N, Phan HTT, Seale AC*, et al.* (2020). Genomic epidemiology of complex, multi-species, plasmid-borne *bla*KPC carbapenemase in Enterobacterales in the UK, 2009-2014. *Antimicrob Agents Chemother*.

Studier FW and Moffatt BA. (1986). Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J Mol Biol* **189**: 113-130.

Suleman M, Jason CR, Bull S and Jones JD. (2024). Ethical argument for establishing good manufacturing practice for phage therapy in the UK. *J Med Ethics*.

Summer EJ, Berry J, Tran TAT, Niu L, Struck DK and Young R. (2007). Rz/Rz1 lysis gene equivalents in phages of Gram-negative hosts. *J Mol Biol* **373**: 1098-1112.

Szermer-Olearnik B and Boratyński J. (2015). Removal of endotoxins from bacteriophage preparations by extraction with organic solvents. *PLoS One* **10**: e0122672.

Tabassum R, Shafique M, Khawaja KA, Alvi IA, Rehman Y, Sheik CS, Abbas Z and Rehman SU. (2018). Complete genome analysis of a *Siphoviridae* phage TSK1 showing biofilm removal potential against *Klebsiella pneumoniae*. *Sci Rep* **8**: 17904.

Tagliaferri TL, Jansen M and Horz HP. (2019). Fighting pathogenic bacteria on two fronts: phages and antibiotics as combined strategy. *Front Cell Infect Microbiol* **9**: 22.

Takeda K, Akimoto C and Kawamukai M. (2001). Effects of the Escherichia coli *sfsA* gene on *mal* genes expression and a DNA binding activity of SfsA. *Biosci Biotechnol Biochem* **65**: 213-217.

Tan D, Zhang Y, Cheng M, Le S, Gu J, Bao J, Qin J, Guo X and Zhu T. (2019). Characterization of *Klebsiella pneumoniae* ST11 isolates and their interactions with lytic phages. *Viruses* **11**: E1080.

Tarkkanen AM, Allen BL, Williams PE, Kauppi M, Haahtela K, Siitonen A, Ørskov I, Ørskov F, Clegg S and Korhonen TK. (1992). Fimbriation, capsulation, and iron-scavenging systems of *Klebsiella* strains associated with human urinary tract infection. *Infect Immun* **60**: 1187-1192.

Teng T, Li Q, Liu Z, *et al.* (2019). Characterization and genome analysis of novel *Klebsiella* phage Henu1 with lytic activity against clinical strains of *Klebsiella pneumoniae*. *Arch Virol* **164**: 2389-2393.

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

Theuretzbacher U. (2017). Global antimicrobial resistance in Gram-negative pathogens and clinical need. *Curr Opin Microbiol* **39**: 106-112.

Thiry D, Passet V, Danis-Wlodarczyk K, *et al.* (2019). New bacteriophages against emerging lineages ST23 and ST258 of *Klebsiella pneumoniae* and efficacy assessment in *Galleria mellonella* larvae. *Viruses* **11**: E411.

Thompson AM and Bizzarro MJ. (2008). Necrotizing enterocolitis in newborns: pathogenesis, prevention and management. *Drugs* **68**: 1227-1238.

Thompson MK, Keithly ME, Sulikowski GA and Armstrong RN. (2015). Diversity in fosfomycin resistance proteins. *Perspect Sci* **4**: 17-23.

Thorvaldsdottir H, Robinson JT and Mesirov JP. (2013). Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform* **14**: 178-192.

Thurgood TL, Sharma R, Call JJ, *et al.* (2020). Genome sequences of 12 phages that infect *Klebsiella pneumoniae*. *Microbiol Resour Announc* **9**: e00024-00020.

Tomás JM, Benedí VJ and Jofre JT. (1987). Identification of the cell surface receptor for FC3-2, FC3-3 and FC3-6 bacteriophages from *Klebsiella pneumoniae*. *FEMS Microbiol Lett* **41**: 223-228.

Tommasi R, Brown DG, Walkup GK, Manchester JI and Miller AA. (2015). ESKAPEing the labyrinth of antibacterial discovery. *Nat Rev Drug Discov* **14**: 529-542.

Tooke CL, Hinchliffe P, Bragginton EC, Colenso CK, Hirvonen VHA, Takebayashi Y and Spencer J. (2019). β-lactamases and β-lactamase inhibitors in the 21st Century. *J Mol Biol* **431**: 3472-3500.

Torres-Barceló C, Arias-Sánchez FI, Vasse M, Ramsayer J, Kaltz O and Hochberg ME. (2014). A window of opportunity to control the bacterial pathogen *Pseudomonas aeruginosa* combining antibiotics and phages. *PLoS One* **9**: e106628.

Townsend EM, Moat J and Jameson E. (2020). CAUTI's next top model - Model dependent *Klebsiella* biofilm inhibition by bacteriophages and antimicrobials. *Biofilm* **2**: 100038.

Townsend EM, Kelly L, Gannon L, *et al.* (2021). Isolation and characterization of *Klebsiella* phages for phage therapy. *Phage (New Rochelle)* **2**: 26-42.

Tsai YK, Fung CP, Lin JC, Chen JH, Chang FY, Chen TL and Siu LK. (2011). *Klebsiella pneumoniae* outer membrane porins OmpK35 and OmpK36 play roles in both antimicrobial resistance and virulence. *Antimicrob Agents Chemother* **55**: 1485-1493.

Turner AK, Eckert SE, Turner DJ, Yasir M, Webber MA, Charles IG, Parkhill J and Wain J. (2020). A whole-genome screen identifies *Salmonella enterica* serovar Typhi genes involved in fluoroquinolone susceptibility. *J Antimicrob Chemother* **75**: 2516-2525.

Turner D, Kropinski AM and Adriaenssens EM. (2021). A roadmap for genome-based phage taxonomy. *Viruses* **13**.

Turner D, Adriaenssens EM, Tolstoy I and Kropinski AM. (2021). Phage annotation guide: guidelines for assembly and high-quality annotation. *Phage (New Rochelle)* **2**: 170-182.

Turner D, Shkoporov AN, Lood C*, et al.* (2023). Abolishment of morphology-based taxa and change to binomial species names: 2022 taxonomy update of the ICTV bacterial viruses subcommittee. *Arch Virol* **168**: 74-82.

Turton J, Davies F, Turton J, Perry C, Payne Z and Pike R. (2019). Hybrid resistance and virulence plasmids in "high-risk" clones of *Klebsiella pneumoniae*, including those carrying *bla*$_{NDM-5}$. *Microorganisms* **7**.

Twort FW. (1915). An investigation on the nature of ultra-microscopic viruses. *Lancet* **2**: 1241-1243.

UK Health Security Agency. (2019). *Annual Epidemiological Commentary: Gram-negative Bacteraemia, MRSA Bacteraemia, MSSA Bacteraemia, and C. difficile Infections, up to and Including Financial Year April 2018 to March 2019*. Agency UHS: London.

Unterhauser K, Pöltl L, Schneditz G*, et al.* (2019). *Klebsiella oxytoca* enterotoxins tilimycin and tilivalline have distinct host DNA-damaging and microtubule-stabilizing activities. *Proc Natl Acad Sci USA* **116**: 3774-3783.

van den Berg DF, van der Steen BA, Costa AR and Brouns SJJ. (2023). Phage tRNAs evade tRNA-targeting host defenses through anticodon loop mutations. *eLife* **12**: e85183.

van Duin J and Tsareva N. (2006). Single-stranded RNA phages. *The Bacteriophages,* (Calendar R, ed.) pp. 175-196. Oxford University Press, Oxford.

van Opijnen T and Camilli A. (2013). Transposon insertion sequencing: a new tool for systems-level analysis of microorganisms. *Nat Rev Microbiol* **11**: 435-442.

Van Twest R and Kropinski AM. (2009). Bacteriophage enrichment from water and soil. *Methods Mol Biol* **501**: 15-21.

Vandenheuvel D, Rombouts S and Adriaenssens EM. (2018). Purification of bacteriophages using anion-exchange chromatography. *Methods Mol Biol* **1681**: 59-69.

Verma V, Harjai K and Chhibber S. (2009). Restricting ciprofloxacin-induced resistant variant formation in biofilm of *Klebsiella pneumoniae* B5055 by complementary bacteriophage treatment. *J Antimicrob Chemother* **64**: 1212-1218.

Vernikos G, Medini D, Riley DR and Tettelin H. (2015). Ten years of pan-genome analyses. *Curr Opin Microbiol* **23**: 148-154.

Vlasblom J, Zuberi K, Rodriguez H*, et al.* (2014). Novel function discovery with GeneMANIA: a new integrated resource for gene function prediction in *Escherichia coli*. *Bioinformatics* **31**: 306-310.

Vogwill T and MacLean RC. (2015). The genetic basis of the fitness costs of antimicrobial resistance: a meta-analysis approach. *Evol Appl* **8**: 284-295.

Walckenaer E, Poirel L, Leflon-Guibout V, Nordmann P and Nicolas-Chanoine MH. (2004). Genetic and biochemical characterization of the chromosomal class A β-lactamases of *Raoultella* (formerly *Klebsiella*) *planticola* and *Raoultella ornithinolytica*. *Antimicrob Agents Chemother* **48**: 305-312.

Waldor MK and Mekalanos JJ. (1996). Lysogenic conversion by a filamentous phage encoding cholera toxin. *Science* **272**: 1910-1914.

Walker BJ, Abeel T, Shea T*, et al.* (2014). Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* **9**: e112963.

Walker KA and Miller VL. (2020). The intersection of capsule gene expression, hypermucoviscosity and hypervirulence in Klebsiella pneumoniae. *Curr Opin Microbiol* **54**: 95-102.

Walker KA, Treat LP, Sepúlveda VE and Miller VL. (2020). The small protein RmpD drives hypermucoviscosity in *Klebsiella pneumoniae. mBio* **11**: e01750-01720.

Walker PJ, Siddell SG, Lefkowitz EJ*, et al.* (2021). Changes to virus taxonomy and to the International Code of Virus Classification and Nomenclature ratified by the International Committee on Taxonomy of Viruses (2021). *Arch Virol* **166**: 2633-2648.

Wang C, Li P, Niu W*, et al.* (2019). Protective and therapeutic application of the depolymerase derived from a novel KN1 genotype of *Klebsiella pneumoniae* bacteriophage in mice. *Res Microbiol* **170**: 156-164.

Wang P, Xiao Y, Gao D, Long Y and Xie Z. (2023). The gene *paaZ* of the phenylacetic acid (PAA) catabolic pathway branching point and *ech* outside the PAA catabolon gene cluster are synergistically involved in the biosynthesis of the iron scavenger 7-hydroxytropolone in Pseudomonas donghuensis HYS. *Int J Mol Sci* **24**.

Wannigama DL, Sithu Shein AM, Hurst C*, et al.* (2023). Ca-EDTA restores the activity of ceftazidime-avibactam or aztreonam against carbapenemase-producing *Klebsiella pneumoniae* infections. *iScience* **26**: 107215.

Warner DM, Yang Q, Duval V, Chen M, Xu Y and Levy SB. (2013). Involvement of MarR and YedS in carbapenem resistance in a clinical isolate of *Escherichia coli* from China. *Antimicrob Agents Chemother* **57**: 1935-1937.

Weber-Dąbrowska B, Żaczek M, Łobocka M*, et al.* (2023). Characteristics of environmental *Klebsiella pneumoniae* and *Klebsiella oxytoca* bacteriophages and their therapeutic applications. *Pharmaceutics* **15**: 434.

Weber BS, De Jong AM, Guo ABY, Dharavath S, French S, Fiebig-Comyn AA, Coombes BK, Magolan J and Brown ED. (2020). Genetic and chemical screening in human blood serum reveals unique antibacterial targets and compounds against *Klebsiella pneumoniae*. *Cell Rep* **32**.

Westra ER, van Houte S, Oyesiku-Blakemore S, Makin B, Broniewski JM, Best A, Bondy-Denomy J, Davidson A, Boots M and Buckling A. (2015). Parasite exposure drives selective evolution of constitutive versus inducible defense. *Curr Biol* **25**: 1043-1049.

Whitfield C and Lam M. (1986). Characterisation of coliphage K30, a bacteriophage specific for *Escherichia coli* capsular serotype K30. *FEMS Microbiology Letters* **37**: 351-355.

WHO. (2017). *Prioritization of Pathogens to Guide Discovery, Research and Development of New Antibiotics for Drug-resistant Bacterial Infections, Including Tuberculosis*. WHO: Geneva.

Wick R and Holt K. (2021). Benchmarking of long-read assemblers for prokaryote whole genome sequencing. *F1000Research* **8**.

Wick RR and Holt KE. (2022). Polypolish: short-read polishing of long-read bacterial genome assemblies. *PLoS Comput Biol* **18**: e1009802.

Wick RR, Judd LM and Holt KE. (2023). Assembling the perfect bacterial genome using Oxford Nanopore and Illumina sequencing. *PLoS Comput Biol* **19**: e1010905.

Wick RR, Judd LM, Gorrie CL and Holt KE. (2017). Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput Biol* **13**: e1005595.

Wick RR, Judd LM, Gorrie CL and Holt KE. (2017). Completing bacterial genome assemblies with multiplex MinION sequencing. *Microbial Genomics* **3**: e000132.

Wick RR, Heinz E, Holt KE and Wyres KL. (2018). Kaptive Web: user-friendly capsule and lipopolysaccharide serotype prediction for *Klebsiella* genomes. *J Clin Microbiol* **56**.

Wick RR, Judd LM, Cerdeira LT, Hawkey J, Méric G, Vezina B, Wyres KL and Holt KE. (2021). Trycycler: consensus long-read assemblies for bacterial genomes. *Genome Biol* **22**: 266.

Wiegand I, Hilpert K and Hancock REW. (2008). Agar and broth dilution methods to determine the minimal inhibitory concentration (MIC) of antimicrobial substances. *Nat Protoc* **3**: 163-175.

Wind CM, de Vries HJC and van Dam AP. (2015). Determination of in vitro synergy for dual antimicrobial therapy against resistant *Neisseria gonorrhoeae* using Etest and agar dilution. *Int J Antimicrob Agents* **45**: 305-308.

Wong JLC, Romano M, Kerry LE, Kwong HS, Low WW, Brett SJ, Clements A, Beis K and Frankel G. (2019). OmpK36-mediated carbapenem resistance attenuates ST258 *Klebsiella pneumoniae* in vivo. *Nat Commun* **10**: 3957.

Wong JLC, David S, Sanchez-Garrido J*, et al.* (2022). Recurrent emergence of *Klebsiella pneumoniae* carbapenem resistance mediated by an inhibitory *ompK36* mRNA secondary structure. *Proc Natl Acad Sci USA* **119**: e2203593119.

Wood DE, Lu J and Langmead B. (2019). Improved metagenomic analysis with Kraken 2. *Genome Biol* **20**: 257-269.

Woodford N, Zhang J, Warner M, Kaufmann ME, Matos J, Macdonald A, Brudney D, Sompolinsky D, Navon-Venezia S and Livermore DM. (2008). Arrival of *Klebsiella pneumoniae* producing KPC carbapenemase in the United Kingdom. *J Antimicrob Chemother* **62**: 1261-1264.

Wright C and Wykes M. (2020). Medaka. Available from: https://github.com/nanoporetech/medaka. (Accessed: 22 Sep 2022).

Wright RCT, Friman VP, Smith MCM and Brockhurst MA. (2019). Resistance evolution against phage combinations depends on the timing and order of exposure. *mBio* **10**: e01652-01619.

Wu Q and Stewart V. (1998). NasFED proteins mediate assimilatory nitrate and nitrite transport in *Klebsiella oxytoca* (*pneumoniae*) M5al. *J Bacteriol* **180**: 1311-1322.

Wu Y, Wang R, Xu M, Liu Y, Zhu X, Qiu J, Liu Q, He P and Li Q. (2019). A novel polysaccharide depolymerase encoded by the phage SH-KP152226 confers specific activity against multidrug-resistant *Klebsiella pneumoniae* via biofilm degradation. *Front Microbiol* **10**: 2768.

Wyres KL, Lam MMC and Holt KE. (2020). Population genomics of *Klebsiella pneumoniae*. *Nat Rev Microbiol* **18**: 344-359.

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

Wyres KL, Hawkey J, Hetland MAK, Fostervold A, Wick RR, Judd LM, Hamidian M, Howden BP, Löhr IH and Holt KE. (2019). Emergence and rapid global dissemination of CTX-M-15-associated *Klebsiella pneumoniae* strain ST307. *J Antimicrob Chemother* **74**: 577-581.

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

Wyres KL, Nguyen TNT, Lam MMC*, et al.* (2020). Genomic surveillance for hypervirulence and multi-drug resistance in invasive *Klebsiella pneumoniae* from South and Southeast Asia. *Genome Med* **12**: 11.

Yang F, Jiang Y-L, Zhang J-T*, et al.* (2023). Fine structure and assembly pattern of a minimal myophage Pam3. *Proc Natl Acad Sci USA* **120**: e2213727120.

Yang JY, Fang W, Miranda-Sanchez F, Brown JM, Kauffman KM, Acevero CM, Bartel DP, Polz MF and Kelly L. (2021). Degradation of host translational machinery drives tRNA acquisition in viruses. *Cell Systems* **12**: 771-779.e775.

Yang SK, Yusoff K, Ajat M, Thomas W, Abushelaibi A, Akseer R, Lim SE and Lai KS. (2019). Disruption of KPC-producing *Klebsiella pneumoniae* membrane via induction of oxidative stress by cinnamon bark (*Cinnamomum verum* J. Presl) essential oil. *PLoS One* **14**: e0214326.

Yano B, Taniguchi I, Gotoh Y, Hayashi T and Nakamura K. (2023). Dynamic changes in Shiga toxin (Stx) 1 transducing phage throughout the evolution of O26:H11 Stx-producing *Escherichia coli*. *Sci Rep* **13**: 4935.

Yasir M, Turner AK, Bastkowski S*, et al.* (2020). TraDIS-Xpress: a high-resolution whole-genome assay identifies novel mechanisms of triclosan action and resistance. *Genome Res* **30**: 239-249.

Yoo S, Lee KM, Kim N, Vu TN, Abadie R and Yong D. (2024). Designing phage cocktails to combat the emergence of bacteriophage-resistant mutants in multidrug-resistant *Klebsiella pneumoniae*. *Microbiol Spectr* **12**: e0125823.

Zaman TU, Alrodayyan M, Albladi M, Aldrees M, Siddique MI, Aljohani S and Balkhy HH. (2018). Clonal diversity and genetic profiling of antibiotic resistance among multidrug/carbapenem-resistant *Klebsiella pneumoniae* isolates from a tertiary care hospital in Saudi Arabia. *BMC Infect Dis* **18**: 205.

Zgurskaya HI, Krishnamoorthy G, Ntreh A and Lu S. (2011). Mechanism and Function of the outer membrane channel TolC in multidrug resistance and physiology of Enterobacteria. *Front Microbiol* **2**: 189.

Zhang S, Cheng Y, Ma J, Wang Y, Chang Z and Fu X. (2019). Degp degrades a wide range of substrate proteins in *Escherichia coli* under stress conditions. *Biochem J* **476**: 3549-3564.

Zhao S, Guo Y, Sheng Q and Shyr Y. (2014). Heatmap3: an improved heatmap package with more powerful and convenient features. *BMC Bioinform* **15**: P16.

Zheng W, Pena A, Low WW, Wong JLC, Frankel G and Egelman EH. (2020). Cryoelectron-microscopic structure of the pKpQIL conjugative pili from carbapenem-resistant *Klebsiella pneumoniae*. *Structure* **28**: 1321-1328.

Žiedaitė G, Daugelavičius R, Bamford JK and Bamford DH. (2005). The holin protein of bacteriophage PRD1 forms a pore for small-molecule and endolysin translocation. *J Bacteriol* **187**: 5397-5405.

Zimin AV and Salzberg SL. (2020). The genome polishing tool POLCA makes fast and accurate corrections in genome assemblies. *PLoS Comput Biol* **16**: e1007981.

10. APPENDICES

APPENDIX 10.1. Metadata for a selection of publicly-available phages downloaded from NCBI for comparative genomics

Phage name	Genus	Accession	Genome size (bp)	Reference
Autographiviridae				
Escherichia phage K30	*Przondovirus*	HM480846	40,940	(Whitfield and Lam, 1986)
Klebsiella phage 2044-307w	*Przondovirus*	MF285615	40,048	-
Klebsiella phage Henu1	*Przondovirus*	MK203841	40,352	(Teng *et al.*, 2019)
Klebsiella phage K5	*Przondovirus*	KR149291	41,698	-
Klebsiella phage K11	*Przondovirus*	EU734173	41,181	(Rudolph *et al.*, 1975)
Klebsiella phage KP32	*Przondovirus*	GQ413937	41,119	(Kęsik-Szeloch *et al.*, 2013)
Klebsiella phage KPN3	*Przondovirus*	MN101227	38,503	-
Klebsiella phage Kund-ULIP47	*Przondovirus*	MK380015	41,397	(Thiry *et al.*, 2019)
Klebsiella phage Kp_Pokalde_002	*Przondovirus*	MT425185	41,816	
Klebsiella phage SH-KP152226	*Przondovirus*	MK903728	41,420	(Wu *et al.*, 2019)
Klebsiella phage vB_KpnP_PRA33	*Przondovirus*	KY652723	40,605	(Labudda *et al.*, 2017)
Klebsiella phage vB_KpnP_BIS33	*Przondovirus*	KY652725	41,697	(Labudda *et al.*, 2017)
Klebsiella phage vB_KpnP_KpV767	*Przondovirus*	KX712070	40,395	-
Klebsiella phage KMI1	*Przondovirus*	MN052874	37,414	-
Klebsiella phage SH-Kp 152410	*Przondovirus*	MG835568	40,945	-
Klebsiella phage K5-2	*Przondovirus*	KY389315	41,116	(Hsieh *et al.*, 2017)
Klebsiella phage K5-4	*Przondovirus*	KY389316	40,163	(Hsieh *et al.*, 2017)
Klebsiella phage vB_Kpn-VAC71	*Przondovirus*	MZ571832	40,388	-
Klebsiella phage vB_KpnP_IME205	*Przondovirus*	KU183006	41,310	(Liu *et al.*, 2020)

Klebsiella phage IME264	*Przondovirus*	OL799328	40,671	-
Salmonella phage phiSG-JL2	*Teetrevirus*	EU547803	38,815	(Kwon *et al.*, 2008)
Citrobacter phage SH1	*Teetrevirus*	KU687347	39,434	(Hamdi *et al.*, 2016)
Klebsiella phage vB_KpnP-VAC1	*Teetrevirus*	MZ428229	39,371	(Bleriot *et al.*, 2022)
Escherichia phage T7	*Tseptimavirus*	V01146	39,937	(Dunn and Studier, 1983)
Enterobacteria phage 13a	*Tseptimavirus*	EU734174	38,841	-
Yersinia phage vB_YenP_AP10	*Apdecimavirus*	KT852574	39,235	-
Yersinia phage Berlin	*Berlinvirus*	AM183667	38,564	-
Drexlerviridae				
Klebsiella phage KPN_N141	*Webervirus*	NC_047841	49,090	-
Klebsiella phage KP36	*Webervirus*	NC_029099	49,797	(Kęsik-Szeloch *et al.*, 2013)
Klebsiella phage NJS2	*Webervirus*	NC_048043	50,132	-
Klebsiella phage P1	*Webervirus*	MZ598515	50,675	-
Klebsiella phage KOX1	*Webervirus*	NC_047825	50,526	(Brown *et al.*, 2017)
Klebsiella phage LF20	*Webervirus*	MW417503	50,107	-
Klebsiella phage P528	*Webervirus*	MW021764	51,895	-
Klebsiella phage Skenny	*Webervirus*	NC_049841	49,935	(Gramer *et al.*, 2019)
Klebsiella phage vB_KpnS_Penguinator	*Webervirus*	MN013087	51,678	(Thurgood *et al.*, 2020)
Klebsiella phage vB_KpnS_ZX2	*Webervirus*	MW722081	51,632	-
Klebsiella phage vB_KpnS_Domnhall	*Webervirus*	NC_049835	54,438	(Thurgood *et al.*, 2020)
Klebsiella phage vB_KpnS_KpV522	*Webervirus*	NC_047784	51,099	-
Klebsiella phage vB_Kpn-VAC113	*Webervirus*	MZ571834	49,568	(Bleriot *et al.*, 2023)
Klebsiella phage vB_Kpn-VAC70	*Webervirus*	MZ571831	49,631	(Bleriot *et al.*, 2023)
Straboviridae				
Escherichia phage T4	*Tequatrovirus*	NC_000866	168,903	(Miller *et al.*, 2003)

Klebsiella phage KMI9	*Slopekvirus*	MN101223	168,074	-
Klebsiella phage KP15	*Slopekvirus*	NC_014036	174,436	(Kęsik-Szeloch *et al.*, 2013)
Klebsiella phage KP27	*Slopekvirus*	NC_020080	174,413	(Kęsik-Szeloch *et al.*, 2013)
Klebsiella phage Miro	*Slopekvirus*	NC_041981	176,055	(Mijalis *et al.*, 2015)
Klebsiella phage phi_KPN_S3	*Slopekvirus*	OQ267591	176,202	(Yoo *et al.*, 2024)
Klebsiella phage PMBT1	*Slopekvirus*	NC_042138	175,206	(Koberg *et al.*, 2017)
Klebsiella phage vB_KoM-Pickle	*Slopekvirus*	LR881145	175,221	(Townsend *et al.*, 2021)
Klebsiella phage vB_KoM-Mild	*Slopekvirus*	LR881147	176,856	(Townsend *et al.*, 2021)
Klebsiella phage UTI-K4	*Jiaodavirus*	OL870319	165,805	(Meile *et al.*, 2023)
Klebsiella phage vB_KaeM_LilPanda	*Jiaodavirus*	OK499980	164,872	-
Klebsiella phage vB_KoM-Flushed	*Jiaodavirus*	LR881144	168,773	(Townsend *et al.*, 2021)
Klebsiella phage vB_KpnM_BovinicusUrsus	*Jiaodavirus*	MW021752	166,829	-
Klebsiella phage vB_KpnM_FRZ284	*Jiaodavirus*	MZ602148	166,376	-
Klebsiella phage JD18	*Jiaodavirus*	NC_028686	166,313	-
Klebsiella phage KP1	*Jiaodavirus*	MG751100	167,989	(Kim *et al.*, 2022)
Klebsiella phage Kpn6N	*Jiaodavirus*	OK631811	163,912	(Weber-Dąbrowska *et al.*, 2023)
Klebsiella phage KPV15	*Jiaodavirus*	NC_055715	167,034	(Aleshkin *et al.*, 2016)
Klebsiella phage vB_KaeM_KaAlpha	*Karamvirus*	MN013084	172,334	-
Klebsiella phage Kp_GWPR59	*Pseudotevenvirus*	OP970827	178,206	-
Unclassified *Caudoviricetes*				
Klebsiella phage 1611E-K2-1	*Jedunavirus*	MG197810	47,797	(Lin *et al.*, 2022)
Klebsiella phage vB_KpnM_15-38_KLPPOU148	*Jedunavirus*	MN689778	49,169	-
Klebsiella phage vB_KpnM_IME346	*Jedunavirus*	MK685667	49,482	(Gao *et al.*, 2022)
Klebsiella phage vB_KpnM_KpV52	*Jedunavirus*	NC_041900	47,405	-
Klebsiella phage VLCpiM12a	*Jedunavirus*	NC_071131	48,833	(Beamud *et al.*, 2023)

Klebsiella phage BUCT_47333	*Jedunavirus*	MZ398021	47,333	-
Klebsiella phage JD001	*Jedunavirus*	NC_020204	48,814	(Cui *et al.*, 2012)
Klebsiella phage MEW1	*Jedunavirus*	MT894004	47,129	-
Klebsiella phage vB_KleM_KB2	*Jedunavirus*	MT757392	48,245	(Peng *et al.*, 2023)
Klebsiella phage vB_KpnM_SCNJ1-Y	Unknown	OQ689083	50,360	(Fang *et al.*, 2023)
Pectobacterium phage PEAT2	*Peatvirus*	NC_044940	48,659	(Kalischuk *et al.*, 2019)
Enterobacter phage ATCEA85	Unclassified siphovirus	MN656993	47,484	(Oh *et al.*, 2020)
Klebsiella phage vB_Kp3	Unclassified siphovirus	ON602766	48,513	-
Klebsiella phage vB_KleS-HSE3	Unclassified siphovirus	MT075871	46,747	(Peng *et al.*, 2020)
Klebsiella phage EKq1	Unclassified siphovirus	OR555718	48,244	(Bird *et al.*, 2024)
Klebsiella phage P61_2	Unclassified siphovirus	OR256024	48,439	-
Klebsiella phage vB_Kpn_K21lambda1	Unclassified siphovirus	OY978848	47,505	(Celia *et al.*, 2024)
Klebsiella phage VLCpiS6a	Unclassified siphovirus	ON602731	46,907	(Beamud *et al.*, 2023)
Pseudomonas phage PSA28	Unclassified siphovirus	MZ089737	48,440	(Nordstrom *et al.*, 2022)
Providencia phage Kokobel2	Unclassified siphovirus	MW145138	45,880	(Rakov *et al.*, 2021)
Salmonella phage PMBT28	Unclassified siphovirus	MG641885	48,113	(Koberg *et al.*, 2018)

APPENDIX 10.2. Summary MLST, virulence factor, AMR, defence system, and prophage data for ten remaining foodborne *Klebsiella* genomes. Genomes were characterised by MLST, capsular type, and O-antigen loci. No genome is in the physical collection (open circles). All genomes contained genes for acrAB (dark orange) and oqxAB (yellow) efflux pumps. Gene clusters for core virulence factor predictions are given as present (dark yellow-green) or incomplete (yellow-green). Accessory virulence factor predictions are given as present (dark pink), incomplete (pink), or absent (light pink). No flagella were predicted in any genome (grey). Heatmap (green) shows the predicted number of acquired AMR determinants according to antibiotic class and the number of plasmid variants. fim, type 1 fimbriae locus; mrk, type 3 fimbriae locus; ent, enterobactin locus; iuc, aerobactin locus; iro, salmochelin locus; ybt, yersiniabactin locus; clb, colibactin locus; rmpA/A2, regulators of mucoid phenotype genes; Chr, chromosomal. Heatmap generated with iTOL.

	MLST	Capsular type	O-antigen	Chr β-lactamase
K. pneumoniae 02387-B4	ST45 (CG45)	KL24	O2v1	SHV-1
K. pneumoniae 0310-357-17A	ST15 (CG15)	KL24	O1v1	SHV-28
K. pneumoniae 0310-357-20A	ST15 (CG15)	KL24	O1v1	SHV-28
K. pneumoniae 0366-2797-39	ST15 (CG15)	KL24	O1v1	SHV-28
K. pneumoniae H-26-MC2393-2397-34	ST15 (CG15)	None	O1v1	SHV-28
K. pneumoniae H-26-MC2393-2397-35	ST15 (CG15)	None	O1v1	SHV-28
K. pneumoniae H-27-MC778-782col37	ST14 (CG14)	None	O1v1	Truncated
K. pneumoniae MC148-152-27	ST290	KL21	O1v1	SHV-1
K. pneumoniae MC1638-1642-73m	ST15 (CG15)	KL24	O1v1	SHV-28
K. pneumoniae MC1968-1972-12mix	ST1383	KL110	O2v1	SHV-119

APPENDIX 10.3. Metadata for all isolated phages.

Phage name	Lab code	Source	Isolation host	Plaque morphology	Notes
Klebsiella phage Bolond	CE1	River Bure, Norwich	*K. michiganensis* M7 21 2 #35	V small T++	
Klebsiella phage Fifoon	CE2	River Bure, Norwich	*K. pneumoniae* M26 18 2 #21 KpnN	V small, T+	
Klebsiella phage Oda	CE3	River Yare, Norwich	*K. michiganensis* M7 21 2 #35	Large, halo	
Klebsiella phage Assynt	CE4	Wastewater A1	*K. pneumoniae* M26 18 2 #21 KpnN	V small, halo	
Klebsiella phage Torridon	CE5	Wastewater A2	*K. pneumoniae* M26 18 2 #21 KpnN	V small, halo	
Klebsiella phage Suilven	CE6	Wastewater A3	*K. pneumoniae* M26 18 2 #21 KpnN	V small, halo	
Klebsiella phage Blaven	CE7	River Tud, Norwich	*K. pneumoniae* M26 18 2 #21 KpnN	V small, T+	
Klebsiella phage Toyotomi	CE8	Wastewater A1	*K. michiganensis* M7 21 2 #35	V large, halo	
Klebsiella phage Mera	CE9	Wastewater A2	*K. michiganensis* M7 21 2 #35	V large, halo	
Klebsiella phage Speegle	CE10	Wastewater A3	*K. michiganensis* M7 21 2 #35	V large, halo	
Klebsiella phage Cornelius	CE11	Wastewater A1	*K. michiganensis* M7 21 2 #35	V large, halo	
Klebsiella phage Tokugawa	CE12	Wastewater A2	*K. michiganensis* M7 21 2 #35	V large, halo	
	CE13	Wastewater A3	*K. michiganensis* M7 21 2 #35	-	Same as CE9
Klebsiella phage SteelHaze	CE14	River Yare, Norwich	*K. michiganensis* M7 21 2 #35	Large, halo	
Klebsiella phage Saitama	CE15	Wastewater B	*K. quasipneumoniae* P057K W	V large, halo	
	CE16	Wastewater B	*K. quasipneumoniae* P057K W	-	Same phage
	CE17	Wastewater B	*K. quasipneumoniae* P057K W	-	
Klebsiella phage Iolaire	CE18	Wastewater C	*K. quasipneumoniae* P057K W	Small, halo	
Klebsiella phage Fioreun	CE19	Wastewater C	*K. quasipneumoniae* P057K W	Small, halo	
Klebsiella phage Emom	CE20	Wastewater B	*K. grimontii* M59 22 8 KoN*	V large, halo	Same phage
	CE21	Wastewater B	*K. grimontii* M59 22 8 KoN*	-	
Klebsiella phage Amrap	CE22	Wastewater C	*K. grimontii* M59 22 8 KoN*	V large, halo	Same phage

	CE23	Wastewater C	*K. grimontii* M59 22 8 KoN*		
Klebsiella phage Kandur	CE24	Wastewater B	*K. grimontii* P038I	V small, halo	
Klebsiella phage NejNej	CE25	Wastewater B	*K. grimontii* P038I	V small, halo	
Klebsiella phage Pigeon	CE26	Wastewater C	*K. grimontii* P038I	V small, halo	
Klebsiella phage Sidders-16	CE27	Wastewater C	*K. grimontii* P038I	V small, halo	
Klebsiella phage Eggie	CE28	Wastewater D	*K. grimontii* P038I	V small, halo	Same phage
	CE29	Wastewater D	*K. grimontii* P038I	-	
Klebsiella phage Whistle	CE30	Wastewater B	*K. variicola* DSM15968	V large, halo	Same phage
	CE31	Wastewater B	*K. variicola* DSM15968	-	
Klebsiella phage DeeWhy	CE32	Wastewater C	*K. variicola* DSM15968	Small, v large halo	Same phage
	CE33	Wastewater C	*K. variicola* DSM15968	-	
Klebsiella phage Dereham	CE34	Wastewater D	*K. variicola* DSM15968	V small, halo	Same phage
	CE35	Wastewater D	*K. variicola* DSM15968	V small, halo	

Wastewater A, wastewater treatment plant located in SE England; A1, raw sewage; A2, aeration feed; A3, mixed liquor; Wastewater B-D, three different wastewater treatment plants located in NW England; *previously misassigned as *K. oxytoca* M59 22 8 KoN. V, very; T, turbidity; +, slightly turbid; ++, very turbid.

APPENDIX 10.4. Distribution of host range breadths of lytic Klebsiella phages. Twenty-five genomically distinct phages were tested against ≥ 34 *Klebsiella* isolates. Number of *Klebsiella* isolates infected (blue) with calculated mean (6.96, SD = 6.88) (grey line) relative to the total number of isolates tested (grey) with calculated mean (36.53, SD = 1.44) (black line).

No. bacterial isolates tested

No. isolates infected   Total no. isolates tested
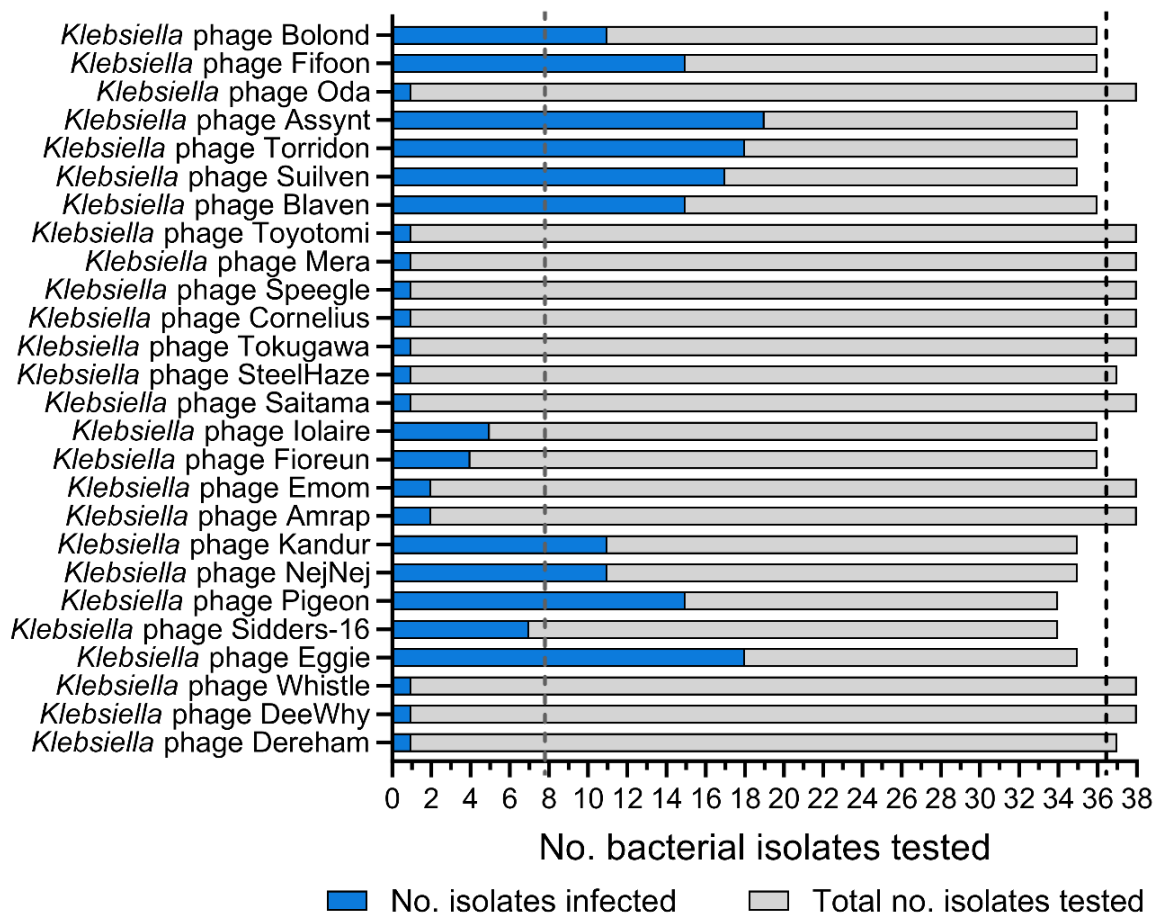
APPENDIX 10.5. Transmission electron microscopy images of Bolond and Oda in combination at a ratio of 1:1. Micrographs were taken at magnifications **(A)** 11kX, **(B)** 13.5kX, **(C)** 28kX, and **(D)** 36kX. Scale bars are given in bottom left.
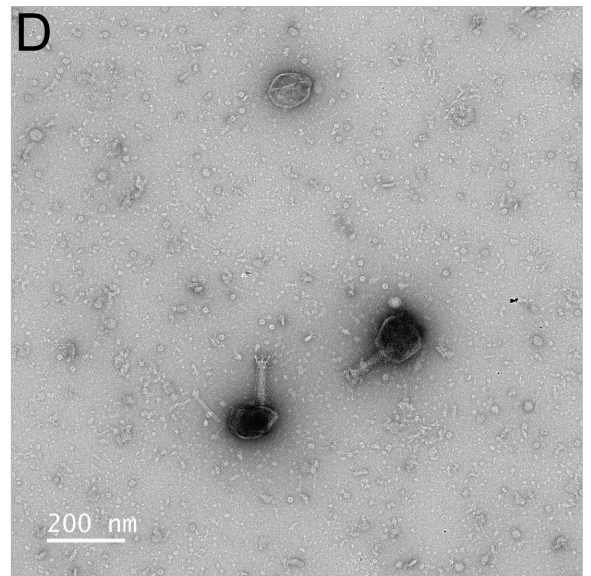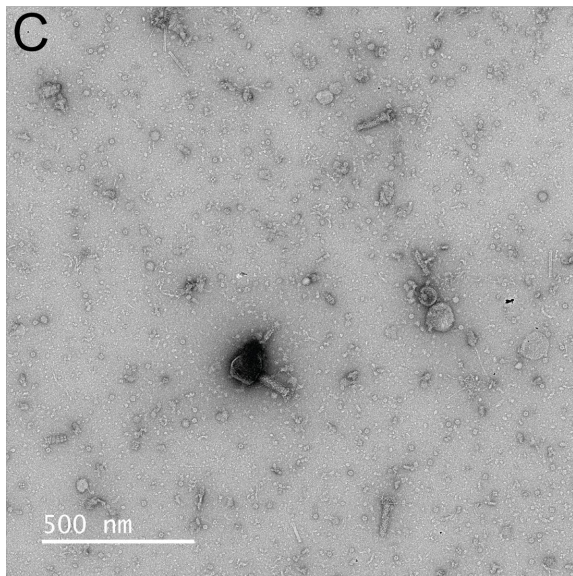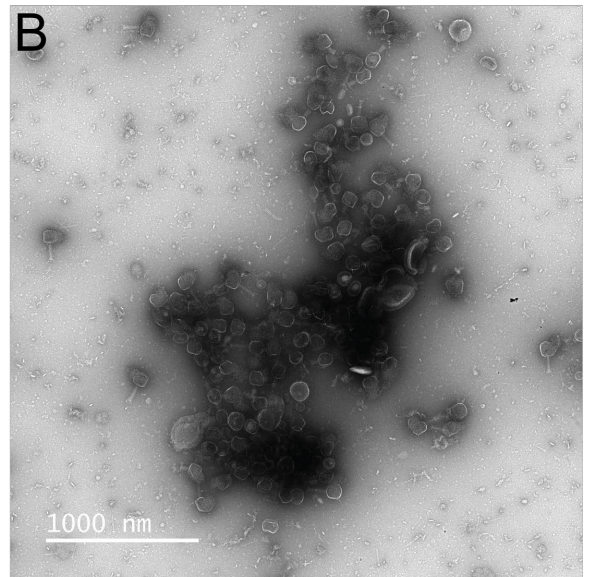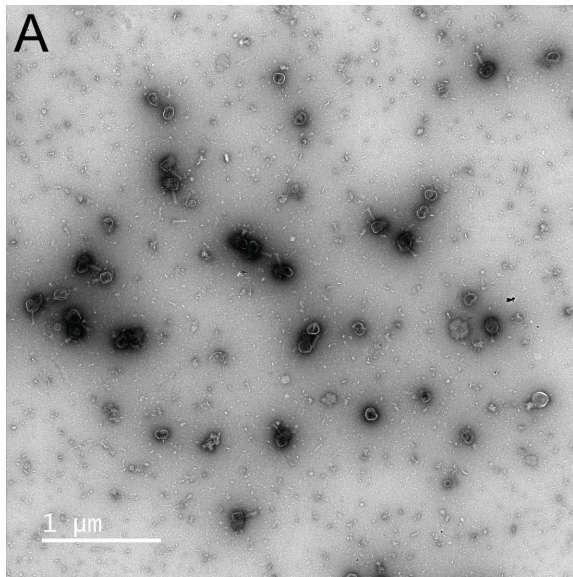
APPENDIX 10.6. Computer-readable data of phage-antibiotic synergy testing with phage Oda.

CHL (µg/mL)

Oda (MOI)	4	2	1	0.5	0.25	0.125	0.0625	0.03125	0.015625	0.007813	0.003906	0
10	0.001	0.000	0.003	-0.002	0.002	-0.003	0.002	0.000	0.001	0.003	0.000	0.000
1	0.001	0.000	0.001	0.000	0.000	-0.003	0.002	-0.004	0.002	0.002	-0.003	0.000
0.1	0.003	0.002	0.001	0.000	0.000	0.002	0.001	-0.001	0.057	-0.001	0.001	0.007
0.01	-0.001	0.003	0.000	-0.003	0.004	0.000	0.000	0.001	0.001	0.001	-0.001	0.001
0.001	0.002	0.001	0.002	0.015	0.000	-0.001	0.003	0.001	0.026	0.001	-0.001	0.002
0.0001	0.001	0.003	0.001	0.066	0.073	0.024	-0.001	0.266	0.007	0.004	0.002	0.004
0.00001	0.001	0.007	0.247	0.000	0.134	0.000	0.390	0.005	0.004	0.140	0.394	0.004
0	0.002	0.011	0.558	0.887	1.096	1.178	1.241	1.237	1.257	1.285	1.195	1.237

CIP (µg/mL)

Oda (MOI)	0.03125	0.015625	0.007813	0.003906	0.001953	0.000977	0.000488	0.000244	0.000122	0.00006	0.000031	0
10	0.001	-0.002	-0.002	0.010	0.000	0.003	-0.002	-0.001	-0.001	0.001	-0.005	0.002
1	0.003	0.011	0.004	-0.003	0.005	-0.001	-0.002	-0.002	0.003	0.001	-0.002	-0.003
0.1	-0.001	-0.003	0.002	0.003	0.001	0.000	-0.001	0.004	0.000	0.002	-0.002	0.001
0.01	0.002	-0.001	0.004	0.003	0.004	0.002	0.005	0.002	0.005	0.001	-0.001	0.002
0.001	0.002	0.005	-0.001	0.005	0.002	0.004	0.065	0.004	0.005	0.001	0.273	-0.001
0.0001	0.005	-0.001	0.002	0.019	0.010	0.002	0.188	0.003	0.002	0.001	0.380	0.000
0.00001	0.006	0.003	-0.001	0.005	0.072	0.506	0.341	0.415	0.010	0.003	0.524	0.003
0	0.007	0.006	0.674	0.975	0.977	0.881	0.908	0.991	1.007	1.088	1.014	0.996

APPENDIX 10.7. Computer-readable data of phage-antibiotic synergy testing with phage Bolond.

Bolond (MOI)	CAR (µg/mL)											
	1024	512	256	128	64	32	16	8	4	2	1	0
10	-0.016	0.256	0.202	0.122	0.364	0.304	0.298	0.268	0.693	0.502	0.740	0.657
1	-0.002	0.219	0.314	0.253	0.347	0.651	0.657	0.332	0.397	0.501	0.359	0.481
0.1	-0.005	0.367	0.336	0.580	0.252	0.148	0.260	0.127	0.144	0.394	0.458	0.368
0.01	0.053	0.337	0.414	0.530	0.450	0.206	0.113	0.223	0.525	0.307	0.687	0.437
0.001	0.058	0.346	0.434	0.996	0.686	0.764	0.696	0.609	0.622	0.643	0.950	0.854
0.0001	-0.018	0.268	0.406	1.033	0.917	1.185	1.140	1.012	1.178	0.974	0.915	0.958
0.00001	-0.109	0.270	0.514	1.138	1.163	1.144	1.129	0.972	1.195	1.019	0.959	1.106
0	-0.014	0.203	0.414	1.133	1.169	1.242	1.159	1.199	1.211	1.164	1.123	1.202

Bolond (MOI)	CTX (µg/mL)											
	0.125	0.0625	0.03125	0.015625	0.007813	0.003906	0.001953	0.000977	0.000488	0.000244	0.000122	0
10	-0.003	-0.004	0.082	0.359	0.162	0.091	0.260	0.068	0.122	0.317	0.392	0.329
1	0.002	-0.004	0.160	0.243	0.179	0.178	0.275	0.330	0.127	0.369	0.188	0.125
0.1	0.012	0.022	0.227	0.260	0.237	0.117	0.072	0.195	0.266	0.115	0.231	0.323
0.01	0.002	0.023	0.472	0.495	0.437	0.181	0.188	0.022	0.092	0.241	0.153	0.163
0.001	0.001	0.040	0.365	0.396	0.543	0.751	0.473	0.428	0.553	0.500	0.399	0.525
0.0001	0.000	0.093	0.095	0.302	0.767	1.070	0.795	0.606	0.704	0.808	0.926	0.977
0.00001	0.010	0.135	0.362	0.277	0.769	1.093	1.106	1.028	1.134	1.025	1.068	1.167
0	0.001	-0.001	0.001	0.232	0.636	1.144	0.998	1.178	1.180	1.180	1.191	1.199

CHL (µg/mL)

Bolond (MOI)	4	2	1	0.5	0.25	0.125	0.0625	0.03125	0.015625	0.007813	0.003906	0
10	0.001	0.002	0.125	0.600	0.293	0.207	0.040	0.091	0.476	0.350	0.492	0.532
1	0.002	0.003	0.032	0.667	0.047	0.015	0.078	0.283	0.036	0.318	0.199	0.023
0.1	0.000	0.002	0.071	0.262	0.205	0.176	0.289	0.200	0.015	0.166	0.374	0.083
0.01	0.002	-0.001	0.308	0.342	0.444	0.098	0.057	0.165	0.365	0.445	0.253	0.412
0.001	0.000	0.002	0.417	0.711	0.217	0.131	0.152	0.246	0.278	0.371	0.327	0.414
0.0001	0.002	0.005	0.510	0.458	0.497	0.307	0.146	0.068	0.084	0.099	0.157	0.118
0.00001	0.003	0.000	0.424	0.926	0.890	0.515	0.348	0.489	0.604	0.784	0.692	0.629
0	0.004	0.011	0.209	0.482	1.191	0.605	1.271	1.222	1.268	1.272	1.283	1.206

CIP (µg/mL)

Bolond (MOI)	0.03125	0.015625	0.007813	0.003906	0.001953	0.000977	0.000488	0.000244	0.000122	0.00006	0.000031	0
10	-0.013	-0.015	0.187	0.177	0.302	0.452	0.494	0.251	0.240	0.462	0.157	0.497
1	-0.011	0.180	0.405	0.255	0.297	0.283	0.363	0.540	0.351	0.491	0.287	0.335
0.1	-0.009	0.137	0.581	0.475	0.256	0.379	0.407	0.263	0.407	0.482	0.239	0.042
0.01	-0.012	0.124	0.708	0.654	0.262	0.309	0.282	0.191	0.419	0.211	0.519	0.339
0.001	-0.010	0.143	0.803	0.985	0.970	0.488	0.419	0.673	0.587	0.374	0.748	0.763
0.0001	-0.009	0.121	0.806	0.788	0.742	1.161	0.978	1.067	1.228	1.164	1.072	1.024
0.00001	-0.015	0.072	0.818	1.041	1.129	1.101	0.865	1.237	1.255	1.187	1.103	1.200
0	-0.011	-0.011	0.713	0.961	1.015	1.037	1.173	1.109	1.204	1.195	1.176	1.212

APPENDIX 10.8. Computer-readable data of phage-antibiotic synergy testing with phage SteelHaze.

CAR (µg/mL)

SteelHaze (MOI)	1024	512	256	128	64	32	16	8	4	2	1	0
10	0.003	-0.002	0.001	0.001	0.003	0.003	0.001	0.002	0.003	0.004	-0.013	0.005
1	0.004	0.002	0.006	0.001	0.007	0.003	0.003	0.006	0.003	0.005	0.006	0.002
0.1	0.005	0.002	0.006	0.003	0.004	0.003	0.001	0.005	0.004	0.002	-0.007	0.001
0.01	0.001	0.005	0.007	0.011	0.013	0.328	0.005	0.010	0.231	0.283	0.016	0.075
0.001	0.002	0.009	0.056	0.044	0.235	0.210	0.157	0.021	0.327	0.453	0.036	0.031
0.0001	0.058	0.220	0.050	0.213	0.340	0.107	0.055	0.247	0.574	0.240	0.130	0.379
0.00001	0.004	0.294	0.174	0.050	1.221	0.937	0.931	1.014	1.346	0.518	1.380	1.012
0	-0.004	0.355	0.321	0.669	1.240	0.983	1.447	1.425	1.429	1.455	1.439	0.984

CTX (µg/mL)

SteelHaze (MOI)	0.125	0.0625	0.03125	0.015625	0.007813	0.003906	0.001953	0.000977	0.000488	0.000244	0.000122	0
10	0.000	-0.004	0.000	0.000	0.000	-0.005	0.000	-0.001	-0.002	-0.004	-0.004	0.002
1	0.002	0.002	-0.001	0.002	0.000	0.023	-0.001	0.000	0.020	-0.001	-0.002	-0.001
0.1	0.002	0.004	-0.001	0.000	0.001	0.000	0.014	0.002	0.000	0.001	0.002	0.003
0.01	0.003	0.001	-0.001	0.004	0.003	0.008	0.112	0.122	0.015	0.005	0.233	0.119
0.001	0.003	0.014	0.010	0.013	0.021	0.191	0.199	0.046	0.124	0.515	0.451	0.013
0.0001	0.003	0.002	0.069	0.184	0.043	0.056	0.158	0.343	0.081	0.236	0.073	0.315
0.00001	0.001	0.001	0.047	0.254	0.607	1.250	0.076	0.946	0.527	0.721	0.920	0.686
0	0.000	0.005	0.378	0.401	0.755	1.352	1.431	1.391	1.467	1.477	1.461	1.226

APPENDIX 10.8 cont.

CHL (µg/mL)

SteelHaze (MOI)	4	2	1	0.5	0.25	0.125	0.0625	0.03125	0.015625	0.007813	0.003906	0
10	0.000	0.000	0.000	0.003	0.001	0.004	0.000	-0.006	0.003	0.040	0.021	-0.003
1	0.001	0.000	0.007	0.001	0.001	0.002	0.003	0.004	0.003	-0.002	-0.002	0.005
0.1	0.000	-0.034	0.002	0.001	0.001	0.003	0.017	0.000	0.435	0.002	0.517	0.007
0.01	0.002	-0.001	-0.001	0.004	0.064	0.248	0.319	0.000	0.305	0.238	0.375	0.343
0.001	-0.002	0.006	-0.001	0.005	0.287	0.095	0.587	0.481	0.354	0.488	0.479	0.183
0.0001	0.001	0.008	-0.003	0.018	0.535	0.299	0.532	0.944	0.282	0.571	0.626	0.135
0.00001	-0.002	0.007	0.574	0.458	1.060	1.159	1.149	1.190	1.215	0.752	1.238	0.669
0	0.001	0.007	0.274	0.902	1.205	1.309	1.276	1.295	1.304	1.316	1.283	1.210

CIP (µg/mL)

SteelHaze (MOI)	0.03125	0.015625	0.007813	0.003906	0.001953	0.000977	0.000488	0.000244	0.000122	0.00006	0.000031	0
10	0.002	0.009	0.003	0.002	-0.003	0.001	-0.002	0.002	0.002	0.000	-0.001	-0.002
1	0.000	0.000	0.002	0.000	-0.002	0.001	-0.002	-0.001	-0.001	0.000	-0.001	0.001
0.1	0.001	-0.004	-0.003	-0.004	0.000	-0.003	-0.001	-0.002	0.001	0.022	0.000	0.011
0.01	-0.002	-0.003	-0.002	-0.002	0.000	0.007	0.003	0.004	0.005	0.051	0.000	0.002
0.001	-0.002	-0.002	-0.001	0.026	0.021	0.017	0.098	0.026	0.078	0.215	0.017	0.232
0.0001	0.001	0.027	0.000	0.053	0.360	0.228	0.227	0.109	0.189	0.036	0.165	0.422
0.00001	-0.002	-0.002	-0.002	0.394	0.712	0.140	0.895	1.205	0.999	0.841	0.882	1.364
0	-0.002	-0.001	-0.001	0.782	1.198	1.271	1.325	1.264	1.397	1.436	1.450	0.826

APPENDIX 10.8 cont.

KAN (µg/mL)

SteelHaze (MOI)	4	2	1	0.5	0.25	0.125	0.0625	0.03125	0.015625	0.00781	0.003906	0
10	0.002	0.001	0.002	0.005	0.002	0.001	0.002	0.003	0.004	0.005	0.006	0.004
1	0.000	-0.003	0.002	0.000	-0.001	0.001	0.002	0.001	0.002	0.003	0.000	0.000
0.1	0.000	0.032	0.009	0.000	0.032	0.010	0.043	0.011	0.003	0.008	0.000	0.005
0.01	0.008	0.064	0.338	0.374	0.144	0.142	0.066	0.211	0.130	0.003	0.011	0.505
0.001	-0.005	0.075	0.700	0.529	0.308	0.187	0.245	0.187	0.457	0.027	0.363	0.234
0.0001	-0.002	0.112	0.164	0.324	0.405	0.058	0.504	0.584	0.149	0.366	0.578	0.267
0.00001	0.003	0.994	0.963	0.334	1.242	0.842	1.155	1.024	0.611	0.360	0.744	0.693
0	0.000	0.923	1.176	1.210	1.235	1.138	1.138	1.163	1.132	1.142	1.150	1.178

MEM (µg/mL)

SteelHaze (MOI)	0.5	0.25	0.125	0.0625	0.03125	0.015625	0.007813	0.003906	0.001953	0.00098	0.000488	0
10	-0.047	-0.024	-0.020	0.503	-0.023	0.541	-0.024	0.001	0.462	-0.023	-0.024	-0.020
1	-0.024	0.416	0.454	0.009	-0.024	-0.018	-0.026	0.260	0.002	0.268	-0.032	-0.022
0.1	-0.024	0.310	-0.025	-0.024	0.160	-0.021	-0.022	-0.021	-0.013	-0.020	0.714	-0.015
0.01	-0.028	-0.022	-0.024	0.738	-0.025	0.013	0.197	-0.022	-0.013	-0.018	-0.022	0.439
0.001	0.481	0.460	0.895	0.502	-0.013	0.301	-0.008	-0.007	0.339	-0.003	0.530	0.503
0.0001	0.065	0.442	0.460	0.604	0.172	0.134	0.536	0.676	0.438	0.057	0.731	0.801
0.00001	0.189	0.595	0.642	1.017	0.877	1.129	0.915	1.175	1.183	0.232	1.040	1.170
0	0.056	1.169	1.215	1.200	1.124	1.165	1.192	1.157	1.188	1.192	1.186	1.201

APPENDIX 10.9. Computer-readable data of phage-antibiotic synergy testing with phage Eggie.

Eggie (MOI)	CAR (µg/mL)											
	1024	512	256	128	64	32	16	8	4	2	1	0
10	-0.011	-0.008	-0.011	-0.010	-0.006	-0.008	-0.009	-0.012	-0.010	-0.009	-0.010	-0.006
1	-0.010	-0.012	-0.009	-0.012	-0.011	-0.009	-0.011	-0.009	-0.008	-0.018	-0.009	-0.009
0.1	-0.009	-0.008	0.037	-0.010	-0.007	-0.010	-0.009	-0.008	-0.008	-0.008	-0.007	-0.007
0.01	-0.009	-0.008	-0.007	-0.005	-0.007	-0.006	-0.010	-0.005	-0.003	0.006	-0.005	-0.007
0.001	0.009	0.003	0.003	0.005	0.017	0.009	0.004	0.004	0.002	0.002	-0.002	-0.003
0.0001	-0.009	0.005	0.069	0.200	0.595	0.434	0.309	0.021	0.525	0.546	0.487	0.554
0.00001	0.003	0.035	0.309	0.771	1.078	0.723	0.679	1.190	0.699	1.021	0.641	0.907
0	0.022	0.152	0.289	0.897	1.197	1.190	1.207	0.621	1.205	1.208	1.189	1.219

Eggie (MOI)	CTX (µg/mL)											
	0.125	0.0625	0.03125	0.015625	0.007813	0.003906	0.001953	0.000977	0.000488	0.000244	0.000122	0
10	-0.004	0.000	-0.003	-0.004	-0.002	-0.003	-0.004	-0.003	0.000	-0.002	0.001	-0.004
1	-0.002	0.001	-0.002	-0.003	-0.003	-0.001	-0.003	0.000	-0.006	0.000	-0.002	0.002
0.1	0.004	0.002	0.006	0.002	0.001	0.001	0.004	0.001	0.000	0.005	0.004	0.000
0.01	0.005	-0.004	0.001	0.001	0.006	0.004	0.000	0.001	-0.001	0.006	0.003	0.003
0.001	0.003	0.000	0.043	0.016	0.016	0.018	0.041	0.017	0.009	0.008	0.016	0.006
0.0001	-0.001	0.002	0.013	0.055	0.039	0.136	0.051	0.201	0.301	0.028	0.012	0.043
0.00001	-0.007	0.001	0.004	0.126	0.287	0.924	0.947	1.121	1.133	0.921	0.818	0.416
0	0.003	0.009	-0.002	0.206	0.805	1.195	1.250	1.218	1.198	1.174	1.112	1.073

APPENDIX 10.9 cont.

CHL (µg/mL)

Eggie (MOI)	4	2	1	0.5	0.25	0.125	0.0625	0.03125	0.015625	0.007813	0.003906	0
10	-0.002	-0.002	-0.001	-0.003	-0.002	0.000	0.004	0.005	0.002	0.000	-0.001	0.001
1	0.001	0.000	0.000	-0.004	-0.003	0.001	0.001	0.001	-0.001	0.000	-0.001	0.003
0.1	0.001	0.004	0.089	0.004	-0.002	0.000	0.000	0.006	0.002	-0.005	-0.003	0.000
0.01	-0.001	0.007	0.304	0.173	0.006	0.008	0.003	0.006	0.018	0.001	0.005	0.003
0.001	0.001	0.007	0.454	0.260	0.304	0.012	0.117	0.283	0.442	0.008	0.649	0.007
0.0001	0.016	0.012	0.590	0.682	0.122	0.251	0.019	0.070	0.007	0.198	0.564	0.091
0.00001	0.000	0.008	0.545	0.708	0.591	0.983	0.959	0.836	1.046	0.617	0.959	0.620
0	-0.001	0.012	0.639	0.967	1.038	1.175	1.223	1.221	1.242	1.257	1.231	1.186

CIP (µg/mL)

Eggie (MOI)	0.03125	0.015625	0.007813	0.003906	0.001953	0.000977	0.000488	0.000244	0.000122	0.00006	0.000031	0
10	-0.005	-0.001	-0.005	-0.006	-0.005	0.033	-0.007	-0.008	-0.006	-0.004	-0.003	-0.003
1	-0.004	-0.004	-0.006	-0.005	-0.006	-0.006	-0.006	-0.004	0.033	-0.006	-0.013	-0.004
0.1	-0.006	-0.005	-0.005	0.020	-0.003	0.030	0.010	-0.004	-0.004	-0.003	0.007	-0.014
0.01	-0.006	-0.005	-0.004	0.000	0.021	0.030	0.039	-0.004	0.000	0.021	-0.002	-0.003
0.001	-0.005	-0.007	0.023	0.229	0.089	0.328	0.013	0.093	0.063	0.177	0.384	0.000
0.0001	-0.004	0.009	0.260	0.322	0.147	0.230	0.317	0.118	0.580	0.279	0.277	0.375
0.00001	-0.007	-0.005	0.748	0.710	0.609	0.867	1.080	0.619	0.784	1.094	0.936	1.064
0	-0.004	-0.004	0.730	0.981	1.095	1.070	1.136	1.200	1.191	1.229	1.260	1.198

APPENDIX 10.10. Phage enumeration by plaque assay of phage Assynt following purification by anion-exchange and FPLC. (A-C) Phage Assynt enumeration of run 1 fractions, including the flow through. The 10% elution peak was contained within fractions P3 A2 and P3 A3; the 40% elution peak was contained within fractions P3 B3, P3 B4, and P3 B5. **(D-G)** Phage Assynt enumeration of run 2 fractions, including the flow through. The 5% elution peak was contained within fractions P3 A2, P3 A3, and P3 A4; the 40% elution peak was contained within fractions P3 B3, P3 B4, P3 B5, P3 C1, and P3 C2.
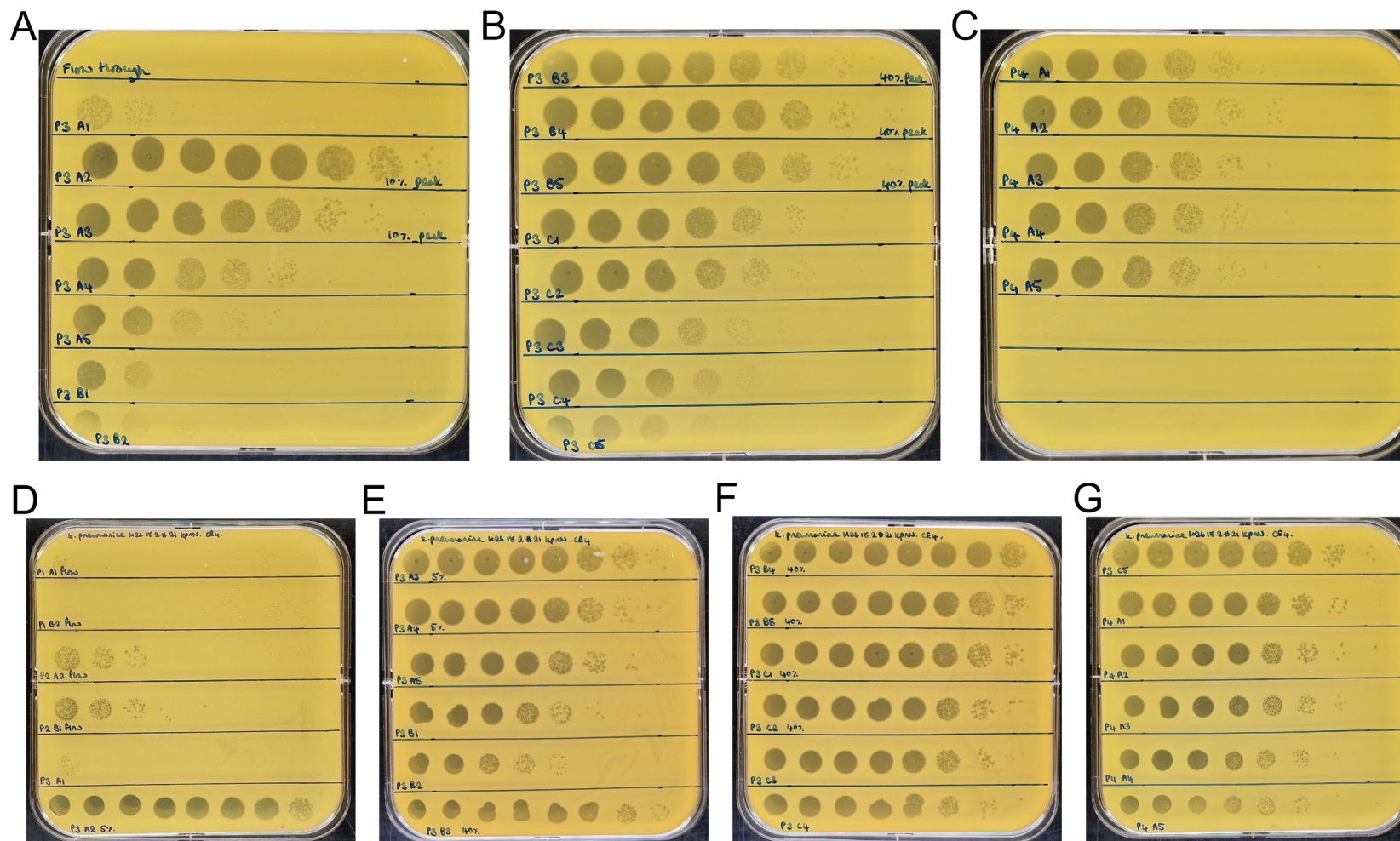
APPENDIX 10.11. Endotoxin data for Klebsiella phage Assynt samples. Fold change relative to crude lysate.

Sample	OD$_{405}$	EU	Dilution	EU/mL	Mean	Fold change
Crude lysate	0.178	0.176	1×10^{-7}	1.76×10^6	-	-
10% 1×10^{-6}	0.956	0.873	1×10^{-6}	8.73×10^5	1.10×10^6	0.63
10% 1×10^{-7}	0.129	0.132	1×10^{-7}	1.32×10^6		
40% 1×10^{-6}	1.200	1.092	1×10^{-6}	1.09×10^6	-	-
40% 1×10^{-7}	0.277	0.265	1×10^{-7}	2.65×10^6	-	1.50

APPENDIX 10.12. Type and number of assembly errors found in short-read-only and long-read only assemblies combined, prior to application of HYPPA.

Phage lab code	Phage name	Assembler	No. repeat region errors	No. homo-polymer errors	No. SNP errors	No. indels[a]	Of which in long-read-only	Of which in short-read-only
CE3	Oda	Canu	2	6	3	1	9	3
CE8[b]	Toyotomi	Flye	2	28	35	1	66	N/A
CE9	Mera	Flye	4	3	2	1	4	6
CE10	Speegle	Flye	2	1	2	1	3	3
CE11	Cornelius	Canu	4	8	1	0	9	4
CE12	Tokugawa	Canu	4	9	0	0	9	4
CE15	Saitama	Canu	4	8	0	0	8	4
CE20	Emom	Canu	4	9	1	0	10	4
CE22	Amrap	Flye	2	2	1	1	3	3
CE30	Whistle	Canu	4	7	0	0	7	4

[a], indels included insertions/deletions separate to both repeat region errors and SNP errors (≥ 2 bp long). [b], no short-read-only assembly for CE8, errors based only on long-read-only assembly prior to HYPPA.

APPENDIX 10.13. Hybrid assembly details for ten przondoviruses

Phage lab code	Phage name	Assembler	No. contigs	Size (bp)
CE3	Oda	N/A	N/A	N/A
CE8	Toyotomi	SPAdes	64	40,626[a]
		Unicycler	6	40,563[a]
CE9	Mera	SPAdes	124	41,329[b]
CE10	Speegle	SPAdes	74	41,270[c]
CE11	Cornelius	N/A	N/A	N/A
CE12	Tokugawa	N/A	N/A	N/A
CE15	Saitama	N/A	N/A	N/A
CE20	Emom	N/A	N/A	N/A
CE22	Amrap	N/A	N/A	N/A
CE30	Whistle	N/A	N/A	N/A

N/A, not performed; [a], genome contained within the first two and largest contigs; [b], genome contained within the first seven and largest contigs; [c], genome contained with the first contig.

APPENDIX 10.14. Alternative long-read-only assembly details for all ten przondoviruses

Phage lab code	Phage name	Assembler	No. contigs	Size (bp)	Size post-polishing (bp)
CE3	Oda	Flye	1	41,116	41,119
CE8	Toyotomi	N/A	N/A	N/A	N/A
CE9	Mera	Canu	1	41,527	40,237
CE10	Speegle	Canu	5	31,413[a]	N/A
CE11	Cornelius	Flye	1	40,487	N/A
CE12	Tokugawa	Flye	1	40,791	40,796
CE15	Saitama	Flye	6	40,551[b]	40,558
CE20	Emom	Flye	1	41,250	N/A
CE22	Amrap	Canu	1	25,100[a]	N/A
CE30	Whistle	Flye	1	40,843	N/A

N/A, not performed; [a], incomplete assembly; [b], genome contained within the first contig.

APPENDIX 10.15. Metadata for failed long-read assemblies for nine phages, where none of the contigs contained the phage that was sequenced.

Phage lab code	Phage name	Assembler	No. of contigs	No. of reads	Read length N50	Longest read (bp)
CE5	Torridon	Flye	93	168,570	270	53,819
CE6	Suilven	Flye	45	1,398	481	21,749
CE7	Blaven	Flye	238	115	68,916	68,916
CE24	Kandur	Flye	10	365,868	258	71,286
CE25*	NejNej	Flye	73	23,516	637	37,626
CE26	Pigeon	Flye	76	48,884	256	52,301
CE27	Sidders-16	Flye	157	77,876	309	85,035
CE28	Eggie	Flye	107	24,709	376	30,529
CE34	Dereham	Flye	57	489	686	12,964

*, phage NejNej long-read-only and subsequent HYPPA assemblies failed, but hybrid assembly with Unicycler was successful.

APPENDIX 10.16. Summary statistics of *K. michiganensis* TraDIS mutant library challenge with phages Oda and SteelHaze (run 1).

Sample	Time point	MOI	Total reads	Reads matched	% matched	Reads mapped	% mapped	Total seq length/total UIS
No phage control	1:30 h	N/A	17,507,727	14,874,752	84.96	174,198	1.171	6332
No phage control	3:00 h	N/A	7,080,487	6,013,774	84.93	70,099	1.166	4284
Phage Oda	1:30 h	10	1,089	1,007	92.47	5	0.497	10
Phage Oda	3:00 h	10	10,734,176	9,665,182	90.04	32,267	0.334	2280
Phage SteelHaze	1:30 h	10	9,033,717	8,315,222	92.05	75,978	0.914	4452
Phage SteelHaze	3:00 h	10	6,992,882	6,371,882	91.12	203,286	3.190	3086

Seq, sequence; UIS, unique insertion sites; h, hours

APPENDIX 10.17. Summary statistics of *K. michiganensis* TraDIS mutant library challenge with phages Bolond, Fifoon, SteelHaze and Eggie (run 2), and phage Oda (run 3). Two independent replicates were performed.

Sample	Time point	MOI	Total reads	Reads matched	% matched	Reads mapped	% mapped	Total seq length/total UIS
Run 2								
No phage control	1:30 h	N/A	22,674,266	20,357,492	89.78	61,436	0.302	4,180
No phage control	1:30 h	N/A	10,253,023	8,915,220	86.95	19,238	0.216	2,802
No phage control	3:00 h	N/A	1,410,307	1,263,734	89.61	4,079	0.323	2,000
No phage control	3:00 h	N/A	10,101,469	8,797,950	87.10	22,821	0.260	3,224
Phage Bolond	1:30 h	1	989,779	8,23,921	83.24	1,100	0.134	1,088
Phage Bolond	1:30 h	1	1,369,723	1,117,296	81.57	1,006	0.090	968
Phage Bolond	3:00 h	1	1,956,720	1,607,548	82.16	1,425	0.089	1,208
Phage Bolond	3:00 h	1	6,707	5,840	87.07	12	0.205	24
Phage Fifoon	1:30 h	1	1,255	1,165	92.83	1	0.086	2
Phage Fifoon	1:30 h	1	1,840	1,758	95.54	3	0.171	4

Phage Fifoon	3:00 h	1	15,698	15,040	95.81	25	0.166	48
Phage Fifoon	3:00 h	1	325	287	88.31	1	0.348	2
Phage SteelHaze	1:30 h	1	253	216	85.38	0	0	0
Phage SteelHaze	1:30 h	1	nd	nd	nd	nd	nd	nd
Phage SteelHaze	3:00 h	1	3,079	2,540	82.49	11	0.433	16
Phage SteelHaze	3:00 h	1	29,3049	16,0805	54.87	910	0.566	308
Phage Eggie	1:30 h	1	188	164	87.23	1	0.610	2
Phage Eggie	1:30 h	1	1,636,037	1,508,924	92.23	3,166	0.210	1,328
Phage Eggie	3:00 h	1	3,464,381	3,119,658	90.05	7,696	0.245	1,902
Phage Eggie	3:00 h	1	3,415,519	3,244,342	94.99	7,308	0.225	1,896
Run 3								
No phage control	0:15 h	N/A	170,451	100,445	58.93	404	0.402	542
No phage control	0:15 h	N/A	318,010	189,205	59.50	745	0.394	800
No phage control	0:30 h	N/A	15,771,701	14,682,455	93.09	40,727	0.277	4,210
No phage control	0:30 h	N/A	5,144,487	4,655,775	90.50	18,455	0.396	2,852
No phage control	1:00 h	N/A	26,470,686	24,472,333	92.45	68,293	0.279	4,790
No phage control	1:00 h	N/A	9,884,415	9,376,517	94.86	27,939	0.298	3,158
Phage Oda	0:15 h	0.1	8,574,223	7,389,957	86.19	23,304	0.315	3,548
Phage Oda	0:15 h	0.1	4,156,455	3,623,779	87.18	9,711	0.268	2,480
Phage Oda	0:30 h	0.1	273,638	164,453	60.10	628	0.382	766
Phage Oda	0:30 h	0.1	3,697,444	3,367,156	91.07	8,502	0.252	2,538
Phage Oda	1:00 h	0.1	8,543,659	7,833,616	91.69	7,983	0.102	2,058
Phage Oda	1:00 h	0.1	5,353,093	5,005,877	93.51	3,244	0.065	1,210
Phage Oda	0:15 h	0.01	9,843,750	8,788,932	89.28	46,445	0.528	4,428

Phage Oda	0:15 h	0.01	1,363,695	1,155,301	84.72	5,984	0.518	2,064
Phage Oda	0:30 h	0.01	9,509,866	8,648,472	90.94	22,729	0.263	3,598
Phage Oda	0:30 h	0.01	7,380,033	6,861,369	92.97	16,375	0.239	2,848
Phage Oda	1:00 h	0.01	7,093,050	6,476,829	91.31	4,601	0.071	2,028
Phage Oda	1:00 h	0.01	7,203,301	6,678,482	92.71	5,035	0.075	2,038

nd, no data; Seq, sequence; UIS, unique insertion sites; h, hours