

Weberviruses are gut-associated phages that infect *Klebsiella* spp.

Samuel J. T. Dawson¹, Preetha Shibu^{2,†}, Sara Garnett¹, Fiona Newberry¹, Thomas C. Brook², Tobi Tijani², Magdalena Kujawska^{3,4,5}, Lindsay J. Hall^{3,4,5,6,7}, Anne L. McCartney¹, David Negus^{1,*}, Lesley Hoyles^{1,*}

¹Department of Biosciences, Nottingham Trent University, Nottingham, NG11 8NS, United Kingdom

²Life Sciences, University of Westminster, London, W1W 6UW, United Kingdom

³Intestinal Microbiome, ZIEL—Institute for Food & Health, Technical University of Munich, Freising 85354, Germany

⁴Microbes, Infection & Microbiomes, College of Medicine & Health, University of Birmingham, Birmingham, B15 2TT, United Kingdom

⁵Institute of Microbiology and Infection, University of Birmingham, Birmingham, B15 2TT, United Kingdom

⁶Gut Microbes & Health, Quadram Institute Bioscience, Norwich Research Park, Norwich, NR4 7UQ, United Kingdom

⁷Norwich Medical School, University of East Anglia, Norwich Research Park, Norwich, NR4 7TJ, United Kingdom

*Corresponding authors. David Negus, Department of Biosciences, Nottingham Trent University, Nottingham, NG11 8NS, United Kingdom.

E-mail: david.negus@ntu.ac.uk; Lesley Hoyles, Department of Biosciences, Nottingham Trent University, Nottingham, NG11 8NS, United Kingdom.

E-mail: lesley.hoyles@ntu.ac.uk

[†]Present address: Berkshire and Surrey Pathology Services, Frimley Health NHS Trust, Wexham Park Hospital, Slough, United Kingdom.

Editor: [Vilje-Petri Friman]

Abstract

Weberviruses are bacteriophages (phages) that can infect and lyse clinically relevant, multidrug-resistant (MDR) strains of *Klebsiella*. They are an attractive therapeutic option to tackle *Klebsiella* infections due to their high burst sizes, long shelf life, and associated depolymerases. In this study, we isolated and characterized seven new lytic phages and compared their genomes with those of their closest relatives. Gene-sharing network, ViPTree proteome, and *terL* gene-sequence-based analyses incorporating all publicly available webervirus genomes [$n = 258$ from isolates, $n = 65$ from metagenome-assembled genome (MAG) datasets] confirmed the seven phages as members of the genus *Webervirus* and identified a novel genus (*Defiantjazzvirus*) within the family *Drexlerviridae*. Using our curated database of 265 isolated phage genomes and 65 MAGs ($n = 330$ total), we found that weberviruses are distributed globally and primarily associated with samples originating from the gut: sewage (154/330, 47%), wastewater (83/330, 25%), and human faeces (66/330, 20%). We identified three distinct clusters of potential depolymerases encoded within the 330 genomes. Due to their global distribution, frequency of isolation and lytic activity against the MDR clinical *Klebsiella* strains used in this study, we conclude that weberviruses and their depolymerases show promise for development as therapeutic agents against *Klebsiella* spp.

Keywords: microbiota; comparative genomics; environment; metagenome-assembled genomes; *Klebsiella pneumoniae*

Introduction

Members of the *Klebsiella pneumoniae* species complex are opportunistic pathogens that can cause serious hospital-acquired infections and are major contributors to global deaths associated with antimicrobial resistance (Antimicrobial Resistance Collaborators 2022). Carbapenem-resistant isolates of *K. pneumoniae* are resistant to a range of frontline β -lactam antibiotics (Antimicrobial Resistance Collaborators 2022, Tsang et al. 2024). The difficulty of treating infections caused by such isolates with conventional antibiotics has resulted in the investigation of new therapeutic modalities, including bacteriophages (phages; viruses that infect and often kill bacteria) and their gene products (Herridge et al. 2020). To realize the potential of phage therapy, it is important to comprehensively characterize phages with clinical potential. Previously we isolated *Webervirus* KLPN1 from the caecum of a healthy female, along with its host *K. pneumoniae* subsp. *pneumoniae* L4-FAA5 (Hoyles et al. 2015). In the current study, we successfully identified seven new representatives of the genus *Webervirus* using L4-FAA5 and multidrug-resistant (MDR) clinical isolates of *Klebsiella* spp. as isolation hosts. These hosts included *K.*

pneumoniae PS_misc6, which encodes the carbapenem-degrading metallo- β -lactamase NDM, and *Klebsiella variicola* PS_misc5, a carbapenem-resistant clinical isolate that encodes the class D β -lactamase OXA-48 (Shibu 2019).

As of 19 January 2025, the genus *Webervirus* encompassed 100 different phage species [International Committee on Taxonomy of Viruses (ICTV)]. With the exception of *Webervirus* BUCT705 (isolated on *Stenotrophomonas maltophilia*), all weberviruses described to date have been isolated on *Klebsiella* hosts, and have proven easy to recover from sewage, wastewater and, occasionally, intestinal contents (Herridge et al. 2020). Although *Webervirus* F20 was originally described as being isolated on *Enterobacter aerogenes* (Mishra et al. 2012), this bacterium has subsequently been reclassified as *Klebsiella aerogenes*. Their high burst sizes (~80 pfu/cell with a reported range between 27 and 142 pfu/cell) (Fang et al. 2022, Gilcrease et al. 2023, Zurabov and Zhilenkov 2021, Li et al. 2024, Senhaji-Kacha et al. 2024, Ziller et al. 2024) and long shelf life make weberviruses ideal phages to work with for biotechnological and clinical applications (Herridge et al. 2020, Fang et al. 2022).

Received 7 November 2024; revised 17 April 2025; accepted 17 April 2025

© The Author(s) 2025. Published by Oxford University Press on behalf of FEMS. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Currently, more than 130 different capsule types (K types) have been identified for *K. pneumoniae* by genetic analysis (Follador et al. 2016). Specific *K. pneumoniae* capsule types are strongly associated with virulence. For example, hypervirulent *K. pneumoniae* isolates are typically associated with capsule types K1, K2, K16, K28, K57, and K63 (Mizuta et al. 1983, Kabha et al. 1995, Yu et al. 2008, Lee et al. 2016, Marr and Russo 2019). Additionally, capsule production by *Klebsiella* spp. has been implicated in protection from complement-mediated lysis and is recognized to play an important role in biofilm formation (Alvarez et al. 2000, Jensen et al. 2020). Capsule type has also been shown to be a major determinant of host tropism in *Klebsiella* phages (Beamud et al. 2023). Bacterial capsules are known to prevent phage attachment by masking cell-surface-associated receptor proteins (Scholl et al. 2005, Dunstan et al. 2021). To overcome this physical barrier, phages encode enzymes—frequently referred to as depolymerases—that selectively degrade polysaccharides or polypeptides that comprise the bacterial capsule (Hoyles et al. 2015, Majkowska-Skropek et al. 2016, Dunstan et al. 2021, Pertics et al. 2021, Cai et al. 2023). Weberviruses tend to have narrow host ranges (Hoyles et al. 2015, Pertics et al. 2021). However, our previous (and ongoing) work has suggested that their depolymerases can degrade the capsules of non-host *Klebsiella* spp. (Hoyles et al. 2015). Depolymerase activity is common to weberviruses, and is being actively investigated as a tool to hydrolyse capsules of *Klebsiella* spp. that often hinder or make treatment with antimicrobials difficult (Majkowska-Skropek et al. 2016, Cai et al. 2019, Pertics et al. 2021). For example, the webervirus depolymerase Depo32 has been shown to protect mice from otherwise lethal *K. pneumoniae* infections in a mouse model of disease (Cai et al. 2023). In addition, a webervirus (P39) has recently been used in combination with another lytic phage (P24, *Przondovirus*) to decolonize mice of carbapenem-resistant *K. pneumoniae* (Fang et al. 2022).

Here, we describe our new webervirus phages and their lytic and depolymerase activities against clinical MDR *Klebsiella* spp., and compare their genomes with those of their closest relatives. The increased ease with which metagenome-associated viruses can be interrogated via PhageClouds (Rangel-Pineros et al. 2021) and NCBI also led us to determine whether weberviruses are readily detectable within recent metagenome-derived phage datasets.

Methods

Bacterial strains

Details of all *Klebsiella* strains included in this study are given in Table 1. The antimicrobial resistance profiles of the isolates, determined according to EUCAST guidelines as described previously (Shibu et al. 2021), can be found in [Supplementary Table 1](#).

Generation of sequence data for bacterial isolates

Genomes for clinical strains included in this study were generated as described previously (Shibu et al. 2021). Illumina and Oxford Nanopore Technologies sequence data for *K. pneumoniae* L4-FAA5 were generated by microbesNG (Birmingham, UK) as described previously (Newberry et al. 2023). CheckM2 v0.1.3 (Chklovski et al. 2022) was used to confirm the quality (in terms of completeness and contamination) of all assembled genomes. Kleborate v3.1.2 (Wyres et al. 2016, Lam et al. 2021) was used to assign sequence types (STs), and capsule (K) and lipopolysaccharide (O) types to genomes.

Isolation and propagation of phages

Filter-sterilized sewage samples were screened for phages as described previously (Smith-Zaitlik et al. 2022) using *Klebsiella* strain L4-FAA5, PS_misc5 or PS_misc6 as inoculum (Table 1). Pure phage stocks were prepared from phage-positive samples as described previously (Hoyles et al. 2015).

Isolation of phage DNA

Phages vB_KpnS-KLPN2, vB_KpnS-KLPN3, and vB_KpnS-KLPN4 were precipitated from 100 ml of each lysate as described previously (Hoyles et al. 2015). Phages vB_KvaS-KLPN5, vB_KvaS-KLPN6, vB_KvaS-KLPN7 and vB_KpnS-KLPN8 were concentrated and DNA extracted as described previously (Smith-Zaitlik et al. 2022).

Transmission electron microscopy

Transmission electron micrographs (TEMs) of phages vB_KpnS-KLPN2, vB_KpnS-KLPN3, and vB_KpnS-KLPN4 were generated as described previously (Hoyles et al. 2015). TEMs for phages vB_KvaS-KLPN5, vB_KvaS-KLPN6, vB_KvaS-KLPN7, and vB_KpnS-KLPN8 were generated and analysed as described previously (Smith-Zaitlik et al. 2022).

Phage genome sequencing, assembly, and annotation

Assembled genomes (from Illumina short-read sequences) for phages vB_KpnS-KLPN2, vB_KpnS-KLPN3, and vB_KpnS-KLPN4 were generated by microbesNG (Shibu et al. 2021). For phages vB_KvaS-KLPN5, vB_KvaS-KLPN6, vB_KvaS-KLPN7, and vB_KpnS-KLPN8, sequence data were generated on an Illumina MiSeq at Nottingham Trent University (Smith-Zaitlik et al. 2022). Quality of raw sequence data was assessed using FastQC v0.11.9. Reads had a mean phred score above 30 and no adapter contamination, so data were not trimmed.

All genomes were assembled using SPAdes v3.13.0 (default settings) (Bankevich et al. 2012), and visualized to confirm circularization of genomes using Bandage v0.8.1 (Wick et al. 2015). CheckV v1.0.1 (checkv-db-v1.5; Nayfach et al. 2021a) was used to determine contamination and completeness of the genomes. Genes in all phage genomes included in this study ([Supplementary Table 2](#)) were predicted and annotated using Pharokka v1.6.1 (v1.4.0 databases) (Bouras et al. 2023).

Comparison of webervirus genomes

ViPTree v4.0 (Nishimura et al. 2017) was used to determine whether the seven phage genomes were closely related to previously described double-stranded DNA viruses. Based on our initial findings (not shown), we curated a list of all known webervirus sequences available from NCBI GenBank on 19 January 2025. We also identified unclassified weberviruses and closely related phage in NCBI using the INPHARED database (1 January 2025 dataset; Cook et al. 2021) and vConTACT v2.0 ([Supplementary Table 2](#)).

Identification of weberviruses in metagenomic datasets

We used PhageClouds (Rangel-Pineros et al. 2021) to identify relatives of weberviruses in metagenome-assembled genome (MAG) datasets. PhageClouds is an online resource that allows researchers to search a reference dataset of ~640 000 phage genomes for phages with genomes related to query sequences.

Table 1. Strains of *Klebsiella* included in this study and their phage infection profiles.

Strain ^a	Source	K:O type ^b	MLST ^c	Infection type ^d									
				KLPN1	KLPN2	KLPN3	KLPN4	KLPN5	KLPN6	KLPN7	KLPN8		
L4-FAA5	Human caecum (Hoyles et al. 2015)	K2:O1ab	ST380	+++++ d	+++++ d	+++++ d	+++++ d	0	0	0	0	0	0
PS_misc2	Groin (Shibu 2019)	K64:O1ab	ST14	0	0	0	0	0	0	0	0	0	0
PS_misc3	Rectum (Shibu 2019)	K52:O13	ST45	0	0	0	0	0	0	0	0	0	0
PS_misc5	Rectum (Shibu 2019)	K81:O13	ST1737-11V	0	0	0	0	+++++ d	+++++ d	0	0	0	0
PS_misc6	Rectum (Shibu 2019)	U:O2a	ST716	0	0	0	0	0	0	0	0	0	++++ d
PS_misc7	Rectum (Shibu 2019)	K30:O1ab	ST294	0	0	0	0	0	0	0	0	0	0
PS_misc8	Perineum (Shibu 2019)	K52:O13	ST45	0	0	0	0	0	0	0	0	0	0
PS_Kpn1	Perineum (Shibu 2019)	K52:O13	ST14-11V	++++	++	++	++	0	0	0	0	0	0
PS_Kpn2	Rectum (Shibu 2019)	K64:O1ab	ST14	++++	+++	+++	+++	0	0	0	0	0	0
PS_Kpn3	Rectum (Shibu 2019)	K64:O3b	ST15-11V	0	0	0	0	0	0	0	0	0	0
PS_Kpn4	Cross-infection (Shibu 2019)	K64:O1ab	ST15	0	0	0	0	0	0	0	0	0	0
PS_Kpn7	Rectum (Shibu 2019)	K2:O1ab	ST14	++	++	++	++	0	0	0	0	0	0
PS_Kpn9	Rectum (Shibu 2019)	K18:O2a	ST515-11V	0	0	0	0	0	0	0	0	0	0
PS_Kpn10	Rectum (Shibu 2019)	K2:O1ab	ST14	+++++	+++	+++	+++	0	0	0	0	0	0
PS_Kpn11	Urine (Shibu 2019)	U:O2afg	ST258	0	0	0	0	0	0	0	0	0	0
PS_Kpn12	Rectum (Shibu 2019)	K15:O4	ST11	0	0	0	0	0	0	0	0	0	0
PS_Kpn13	Rectum (Shibu 2019)	K2:O1ab	ST14	++++	+++	+++	+++	0	0	0	0	0	0
PS_Kpn14	Rectum (Shibu 2019)	K2:O1ab	ST14	+++++	+++	+++	+++	0	0	0	0	0	0
PS_Kpn15	Rectum (Shibu 2019)	K52:O13	ST45	0	0	0	0	0	0	0	0	0	0
PS_Kpn16	Rectum (Shibu 2019)	K17:O1ab	ST101	0	0	0	0	0	0	0	0	0	0
PS_Kpn24	Rectum (Shibu 2019)	K2:O1ab	ST14	++	++	++	++	++	++	++	++	++	0
PS_Kpn25	Rectum (Shibu 2019)	K2:O1ab	ST14	++++	+++	+++	+++	0	0	0	0	0	0
PS_Kpn26	Rectum (Shibu 2019)	K64:O2a	ST147	0	0	0	0	0	0	0	0	0	0
PS_Kpn27	Urine (Shibu 2019)	U:O2afg	ST258	0	0	0	0	0	0	0	0	0	0
PS_Kpn28	Urine (Shibu 2019)	K2:O1ab	ST14	+++++	+++	+++	+++	0	0	0	0	0	0
PS_Kpn29	Urine (Shibu 2019)	K2:O1ab	ST14	++	++	++	++	0	0	0	0	0	0
PS_Kpn30	Mouth (Shibu 2019)	U:O1ab	ST15	0	0	0	0	0	0	0	0	0	0
PS_Kpn31	Perineum (Shibu 2019)	K2:O1ab	ST14	+++++	++	++	++	0	0	0	0	0	0
PS_Kpn32	Drain fluid (Shibu 2019)	K22:O1ab	ST35	0	0	0	0	0	0	0	0	0	0
PS_Kpn33	Urine (Shibu 2019)	K2:O1ab	ST14	+++ (d)	+++ (d)	++	++	0	0	0	0	0	0
PS_Kpn35	Urine (Shibu 2019)	K2:O1ab	ST14	+++ (d)	+++ (d)	++	++	0	0	0	0	0	0
PS_Kpn36	Urine (Shibu 2019)	K24:O2a	ST11	0	0	0	0	0	0	0	0	0	0
PS_Kpn37	Wound (Shibu 2019)	K51:O3b	ST16	0	0	0	0	0	0	0	0	0	0
PS_Kpn38	High vaginal swab (Shibu 2019)	K2:O1ab	ST14	+++	++	+++	+++	0	0	0	0	0	0
PS_Kpn39	Wound (Shibu 2019)	K64:O1ab	ST14	0	0	0	0	0	0	0	0	0	0
PS_Kpn40	Wound (Shibu 2019)	U:O2afg	ST512	0	0	0	0	0	0	0	0	0	0
PS_Kpn41	Leg (Shibu 2019)	U:O1ab	ST15	0	0	0	0	0	0	0	0	0	0

^aAll strains with PS prefix identified as *K. pneumoniae* subsp. *pneumoniae* by average nucleotide identity and phylogenetic analyses against type strains of the genus *Klebsiella*, except for PS_misc5 (*K. varicola* subsp. *varicola*) (Shibu 2019).

^bDetermined using Kleborate. U, untypeable. Full Kleborate dataset available in [Supplementary Table 1](#).

^cMLST, multi-locus sequence type determined using Kleborate.

^dThe appearance of the spot was graded according to Haines et al. (2021): +++++, complete lysis; ++, hazy lysis; +, visible plaques; 0, no visible plaques. We also noted whether depolymerase activity was detected (i.e. formation of halos around plaques): d, depolymerase activity; (d), weak depolymerase activity detected.

The genome of *Webervirus* KLPN1 was searched against the PhageClouds database with a threshold of 0.15, as we had previously looked for relatives of this phage in metagenomic datasets and are interested in gut-associated phage communities (Hoyles et al. 2015). The nucleotide sequences of the relevant phage MAGs (from Gregory et al. 2020, Camarillo-Guerrero et al. 2021, Tisza and Buck 2021) were recovered from the relevant datasets.

Additionally, webervirus MAGs were identified using a search of the NCBI nucleotide database for *Bacteriophage* sp. [search term: (txid38018) AND MAG]; the sequences ($n = 8138$) were filtered for genomes of between 30 Kbp and 60 Kbp in length ($n = 2540$): genes were predicted using Prodigal v.2.6.3 (Hyatt et al. 2010), and the proteomes added to the INPHARED database and analysed using vCONTact2. CheckV was used (as described above) to determine contamination and completeness of the MAG dataset (Supplementary Table 3).

The MAG sequences were analysed using ViPTree v4.0 to confirm their affiliation with the genus *Webervirus*. They were also annotated with Pharokka and included in a vConTACT2 analysis with our curated set of webervirus genomes. The genomes of all weberviruses were compared with one another using taxmyPHAGE v0.3.3, which uses a Python implementation of the VIRIDIC algorithm to calculate intergenomic genomic similarities (Millard et al. 2024). The matrix created from the similarity values was visualized using tidyheatmaps v0.2.1 (Mangiola and Papenfuss 2020).

Phylogenetic relationships among weberviruses

Nucleotide sequences of the large-subunit terminase (*terL*) genes, predicted by Pharokka, were used to create a multiple-sequence alignment (Clustal Omega 1.2.2 implemented in Geneious Prime v2024.0.5; options—group sequences by similarity, 5 representative iterations). This alignment was used to create a bootstrapped (100 replicates) maximum-likelihood tree (PhyML v3.3.20180621, JC69 algorithm).

Distribution of weberviruses

The distribution of weberviruses was determined by identifying the source and geographical location information for the GenBank genomes (including our seven new genomes; Supplementary Table 2) and the MAGs (Supplementary Table 3). Data were aggregated based on isolation source or geographical location, with these latter data visualized using the R package rworldmap v1.3.8 (South 2011).

Host prediction for MAGs

The CRISPR Spacer Database and Exploration Tool (Dion et al. 2021) and HostPhinder 1.1 (Villarroel et al. 2016) were used to predict hosts for the weberviruses recovered from metagenomic datasets. MAGs were also subject to a BLASTN search against the Unified Human Gastrointestinal Genome (UHGG) CRISPR spacer database according to Nayfach et al. (2021b). For this, a BLASTN database was created from 1846 441 spacers from 145 053 CRISPR arrays from 79 735 UHGG genomes (Nayfach et al. 2021b). Spacers were searched against viral genomes using BLASTN from the blast+ package v.2.12.0 (options: -dust=no; -word-size=18); a maximum of one mismatch or gap was allowed over $\geq 95\%$ of the spacer length. iPHoP v1.3.3 (Roux et al. 2023), an automated command-line pipeline for predicting host genus of novel bacteriophages and archaeoviruses based on their genome sequences, was also used to analyse the MAGs.

Identification of potential depolymerases among weberviruses

A BLASTP database was created using webervirus amino acid sequences from experimentally validated webervirus depolymerases, and a BLASTP search was run versus all webervirus genomes. Sequences used to build the BLAST database are available from figshare (doi:10.6084/m9.figshare.28603070). Clustal Omega v1.2.2 alignments were created in Geneious Prime (default settings; 2024.0.5). RAXML v 8.2.11 (-m PROTGAMMABLOSUM62 -f a -x 1 -N 100 -p 1) was used to generate a bootstrapped (100 replicates) maximum-likelihood tree from the multiple-sequence alignment.

Results

Seven new weberviruses lyse a range of clinically relevant *Klebsiella* spp.

Seven phages were isolated on two different strains of *K. pneumoniae* subsp. *pneumoniae* (L4-FAA5-vB_KpnS-KLPN2, vB_KpnS-KLPN3, vB_KpnS-KLPN4; PS_misc6-vB_KpnS-KLPN8) and one strain of *K. variicola* subsp. *variicola* (PS_misc5-vB_KvaS-KLPN5, vB_KvaS-KLPN6, vB_KvaS-KLPN7). All our sewage samples yielded *Klebsiella*-infecting phages. Strain L4-FAA5 (K2: O1ab, ST380) was originally isolated from human caecal effluent along with *Webervirus* KLPN1 (Hoyles et al. 2015), while strains PS_misc5 (K81: O13, ST1737-1LV) and PS_misc6 (untypeable: O2a, ST716) were part of a collection ($n = 36$) of clinical MDR and/or carbapenem-resistant *Klebsiella* isolates currently being used in our laboratory in phage-related and other studies (Shibu 2019) (Supplementary Table 1).

TEM showed the seven phages had a mean capsid diameter of 57.5 nm and a mean tail length of 157.5 nm (Supplementary Fig. A). Host-range analysis showed the seven phages had different infection profiles (Table 1). KLPN1, our original webervirus isolated on *K. pneumoniae* L4-FAA5 (Hoyles et al. 2015), was included in analyses for comparative purposes. KLPN1, vB_KpnS-KLPN2, vB_KpnS-KLPN3, and vB_KpnS-KLPN4 completely lysed some, but not all, clinical isolates of *K. pneumoniae* with capsule: O antigen types K52: O13 and K64: O1ab. K2: O1ab isolates alone were infected by KLPN1 and vB_KpnS-KLPN2 to vB_KpnS-KLPN4, though vB_KpnS-KLPN4 was unable to infect one of the K2: O1ab strains (PS_Kpn13). Only on strain L4-FAA5 (K2: O1ab), isolated from human caecal effluent, was strong depolymerase activity observed with phages KLPN1 and vB_KpnS-KLPN2. Hazy lysis of strain PS_Kpn24 (K2: O1ab) was observed with phages vB_KvaS-KLPN5, vB_KvaS-KLPN6 and vB_KvaS-KLPN7. Phages vB_KvaS-KLPN5 to vB_KvaS-KLPN7 showed strong lytic and depolymerase activity on *K. variicola* PS_misc5 (K81: O13) alone, while vB_KpnS-KLPN8 lysed *K. pneumoniae* PS_misc6 (untypeable capsule: O2a) with depolymerase activity on this host.

Genome-based analyses of publicly available sequence data triples the number of authenticated webervirus genomes

Bandage (data not shown) and CheckV (Supplementary Table 2) analyses confirmed the genomes of vB_KpnS-KLPN2, vB_KpnS-KLPN3, vB_KpnS-KLPN4, vB_KvaS-KLPN5, vB_KvaS-KLPN6, vB_KvaS-KLPN7 and vB_KpnS-KLPN8 were circular and complete. None of the genomes was contaminated. An initial online ViPTree analysis showed vB_KpnS-KLPN2, vB_KpnS-KLPN3, vB_KpnS-KLPN4, vB_KvaS-KLPN5, vB_KvaS-KLPN6, vB_KvaS-KLPN7, and vB_KpnS-KLPN8 belonged to the genus *Webervirus* (data not shown). All publicly available webervirus genomes (available as

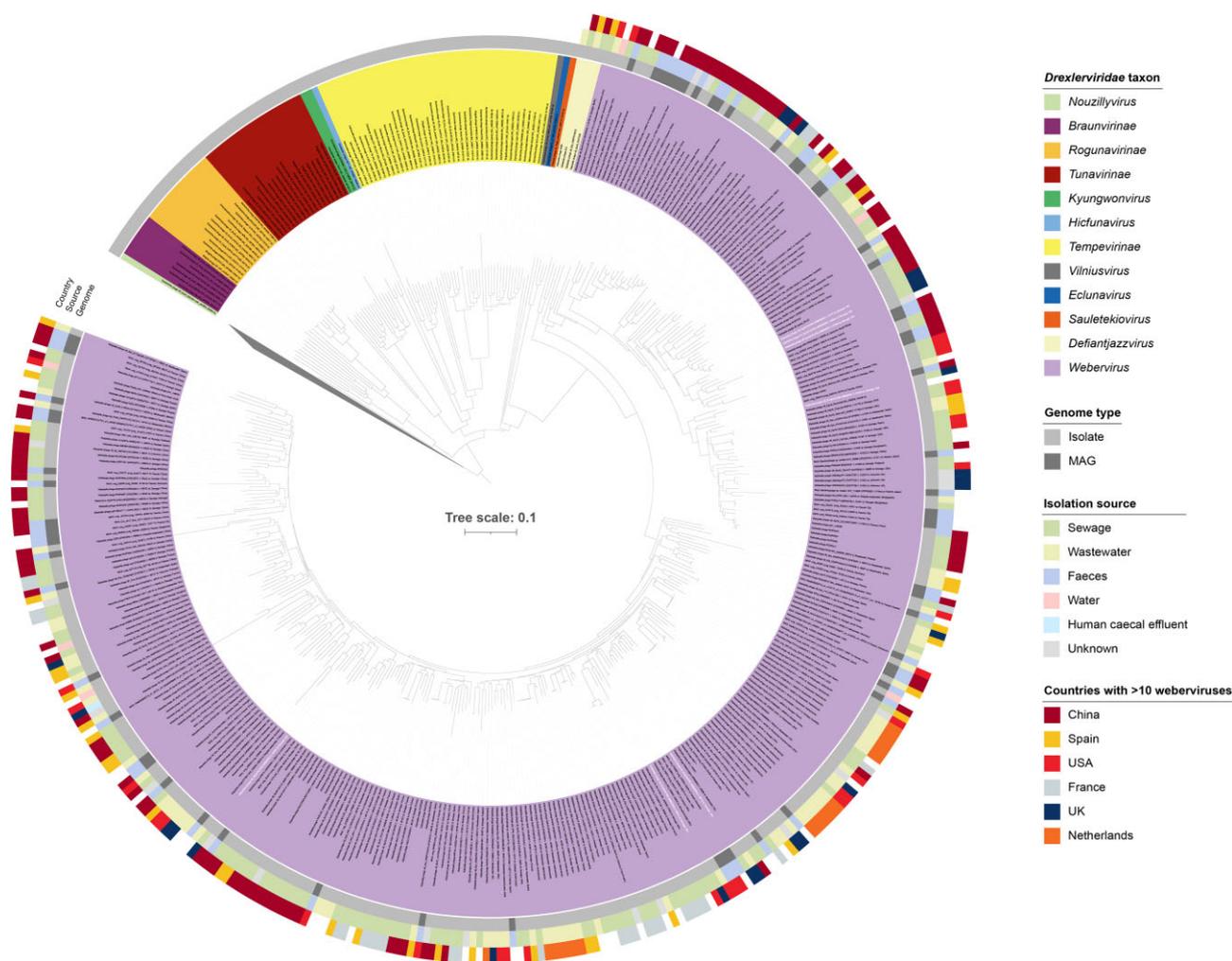


Figure 1. ViPTree-generated phylogenetic analysis of the family *Drexlerviridae*. The genus *Webervirus* is represented by 330 genomes. The names of our seven newly identified weberviruses are shown in white bold text. A potentially novel genus (*Defiantjazzvirus*) was identified during the curation of our dataset. The colours covering the virus names represent taxa within the family *Drexlerviridae*; the outgroup has been collapsed to aid visualization. The tree (ViPTree bionj) was rooted at the midpoint.

of 19 January 2025) were downloaded from GenBank to allow comparison with our newly sequenced phages and for inclusion in the INPHARED vCONTACT2 database if not already included in the 1 January 2025 release. Among the other 264 genomes from phage isolates included in this study, 226 were of high quality, 43 were complete and two were of medium quality; none of these genomes was contaminated.

In addition, we used PhageClouds to identify potential webervirus MAGs. Fifty-four of the PhageClouds hits represented MAGs derived from the Gut Phage Database (GPD) (Camarillo-Guerrero et al. 2021), six were from the Cenote Human Virome Database (CHVD) (Tisza and Buck 2021), and two were from the Gut Virome Database (GVD) (Gregory et al. 2020). MAG Ma_2019_SRR413710_NODE_378_length_50715_cov_48.086538 from the GVD was identical to uvig_330395 from the GPD (Camarillo-Guerrero et al. 2021) so was removed from further analyses [PhageClouds scores identical, 100% pairwise identity as assessed using VIRIDIC; an unsurprising finding as both MAGs are derived from the same dataset (Ma et al. 2018)]. Similarly, two MAGs from the GPD were also found to be identical: uvig_314355 and uvig_315584 were high-quality genomes both derived from the same four samples [SRR1952259, SRR1162648, SRR1162662, SRR1162654 (Tisza and Buck 2021)]; only uvig_314355 was re-

tained for further analyses. Our inclusion of NCBI genomes listed as Bacteriophage sp. in a vCONTACT2 analysis with the INPHARED database identified a further five potential webervirus MAGs recovered from faecal samples in Japan (Nishijima et al. 2022). In total, our dataset included 65 MAGs. The MAGs ranged from 10230 to 55276 nt (mean 42392 nt) in length (Supplementary Table 3). Forty-seven of the 65 MAGs were determined to be complete or of high-quality (CheckV). Eight were of medium-quality and 10 were low quality, representing genome fragments (Supplementary Table 3). None of the MAGs was contaminated.

In addition to the 100 recognized weberviruses included in the ICTV and our seven new weberviruses, we identified 158 more webervirus genomes in NCBI and 65 webervirus MAGs. The 265 weberviruses isolated on bacteria mostly infected *K. pneumoniae* (Supplementary Fig. B). A ViPTree analysis confirmed the affiliation of our 330 genomes with the genus *Webervirus* (Fig. 1). The webervirus genomes often clustered based on geographical origin, irrespective of whether they came from phage isolates or MAGs (Fig. 1).

The monophyletic nature of the genus *Webervirus* was confirmed by phylogenetic analysis of *terL* gene sequences (99% bootstrap support; Fig. 2A). A gene-sharing network was created with

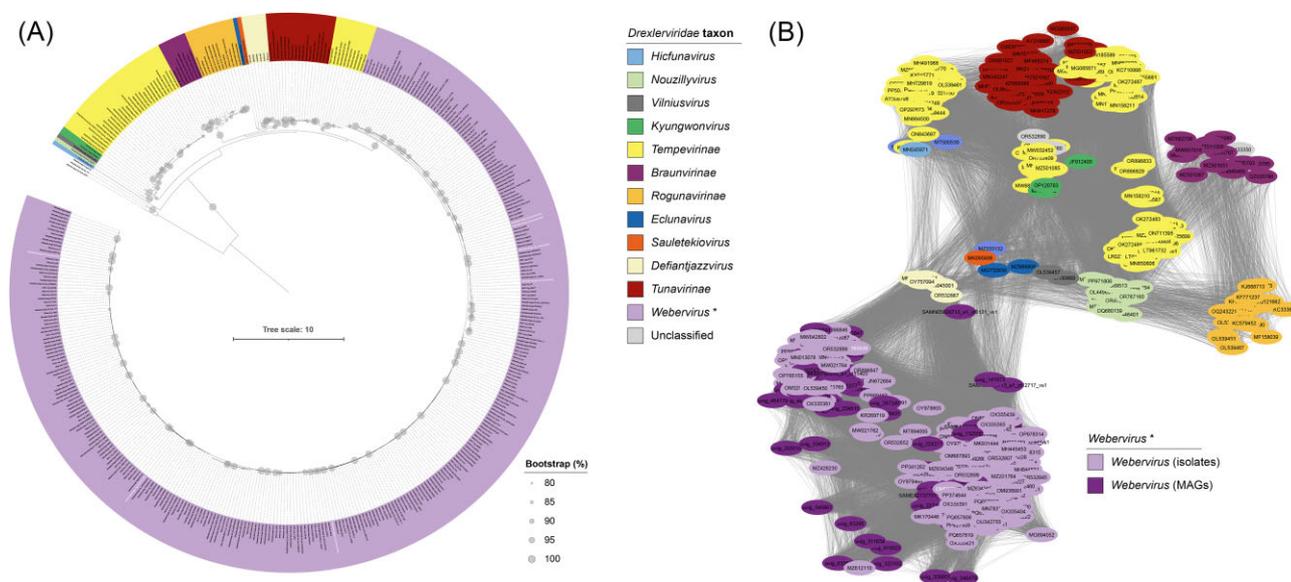


Figure 2. Further analyses of Drexleviridae sequence data. (A) Phylogenetic relationships (maximum-likelihood tree) of members of the family Drexleviridae based on analysis of large-subunit terminase (*terL*) nucleotide sequences encoded in phage genomes. Bootstrap values are expressed as a percentage of 100 replications; scale bar, mean number of nucleotide substitutions per position; the tree is rooted at the midpoint. (B) Gene-network-based analysis of proteomic data for members of the genus *Webervirus* and their nearest relatives. Full network shown in Supplementary Fig. C. (A and B) The legend shown applies to both figures, with isolate and MAG proteomes differentiated in (B). Names of our seven newly identified weberviruses are shown in bold white text.

all webervirus genomes included in this study (Supplementary Table 2 and Supplementary Table 3) using vConTACT v2.0 (Bolduc et al. 2017, Bin Jang et al. 2019) and the INPHARED database (Supplementary Fig. C). The network was filtered based on first and second neighbours of webervirus proteomes (Supplementary Fig. C, Fig. 2B). The vConTACT-based analysis confirmed findings from the ViPTree- and *terL*-based analyses with respect to affiliation of weberviruses included in this study.

VIRIDIC analysis split the weberviruses into eight different clusters at the genus level, with most weberviruses affiliated with Cluster 1 (Supplementary Fig. D, Supplementary Table 4). Clusters 3 (uvig_338 855, uvig_63 295), 4 (uvig_346 479, uvig_474 523), 5 (SAMN05826713_a1_ct6131_vs1), 6 (uvig_63 387), 7 (uvig_340 901), 8 (uvig_334 913), and 9 (SAMN05826713_a1_ct12717_vs 1) were all associated with low-quality MAGs (Supplementary Table 3). MAGs in these clusters shared <70% identity with Cluster 1 phages (isolate and MAG genomes). The only other low-quality MAG included in the analysis (uvig_311 634) was affiliated with Cluster 1 phages, sharing 33%–72% identity with them and highest similarity with a MAG (uvig_141 073) in this cluster (Supplementary Table 4).

Identification of a novel genus within the family Drexleviridae

Our ViPTree analysis also identified a potential novel genus (referred to as *Defiantjazzvirus*) comprising six representatives within the family Drexleviridae and closely related to the genus *Webervirus* (Fig. 1). Analysis of *terL* gene sequences showed this genus to be monophyletic (97% bootstrap support; Fig. 2A). vConTACT-based analysis demonstrated that the six genomes associated with *Defiantjazzvirus* clustered together but separately from all other phage groups included in the analysis (Fig. 2B). VIRIDIC analysis showed *defiantjazzvirus* genomes to share 81%–97% genome identity with one another and 27%–42% identity with members of the genus *Webervirus* (Supplementary Fig. D, Supplementary Table 4). Based on current recommendations, the six genomes (sharing >70%

nucleotide identity across their full-length genomes) represent a novel genus comprising five species (Supplementary Table 4) (Turner et al. 2021). Comparison of the *defiantjazzvirus* genomes with non-webervirus Drexleviridae genomes confirmed the genus *Defiantjazzvirus* represents a novel genus within the family Drexleviridae, with the six *defiantjazzvirus* genomes sharing between 0.2% and 35% genome identity with their closest non-webervirus relatives (Supplementary Table 5). Representatives of the genus *Defiantjazzvirus* infect *K. pneumoniae*, *K. michiganensis*, and *K. oxytoca* (Supplementary Table 2).

Webervirus MAGs are predicted to infect *Klebsiella*

To confirm the MAGs were weberviruses that infected *Klebsiella* spp., we attempted to predict their bacterial hosts. CRISPR spacers can be used to predict hosts of unknown phages, as spacers represent biological records of past phage–bacteria interactions. Each of the seven new phage genomes (nt sequences) we generated was uploaded to CRISPR Spacer Database and Exploration Tool (Dion et al. 2021). None of the phages could be assigned to known hosts using this tool. Using the BLASTN approach of Nayfach et al. (2021b) with the MAG sequences, only SAMEA2737751_a1_ct5309 had sufficient coverage; this MAG had two hits to *Klebsiella* species (*K. pneumoniae* and *K. variicola*). iPhoP predicted hosts for 84/330 of the webervirus genomes included in this study; *Escherichia* was predicted to be the host for 21 of the MAGs and 59 of the isolates at the genome and genus levels (Supplementary Table 6). Only NC_049845.1, OR532813.1, OR532891, PQ337355, and PQ519586—all representing isolated phages (Supplementary Table 2)—were predicted to have a *Klebsiella* host at the genus level. HostPhinder 1.1 (Villarreal et al. 2016) was able to predict hosts for our KLPN phages, with all assigned to *Klebsiella pneumoniae*. Consequently, this tool was used to predict hosts for the *Webervirus* MAGs (Supplementary Table 7). All were predicted to infect *Klebsiella*.

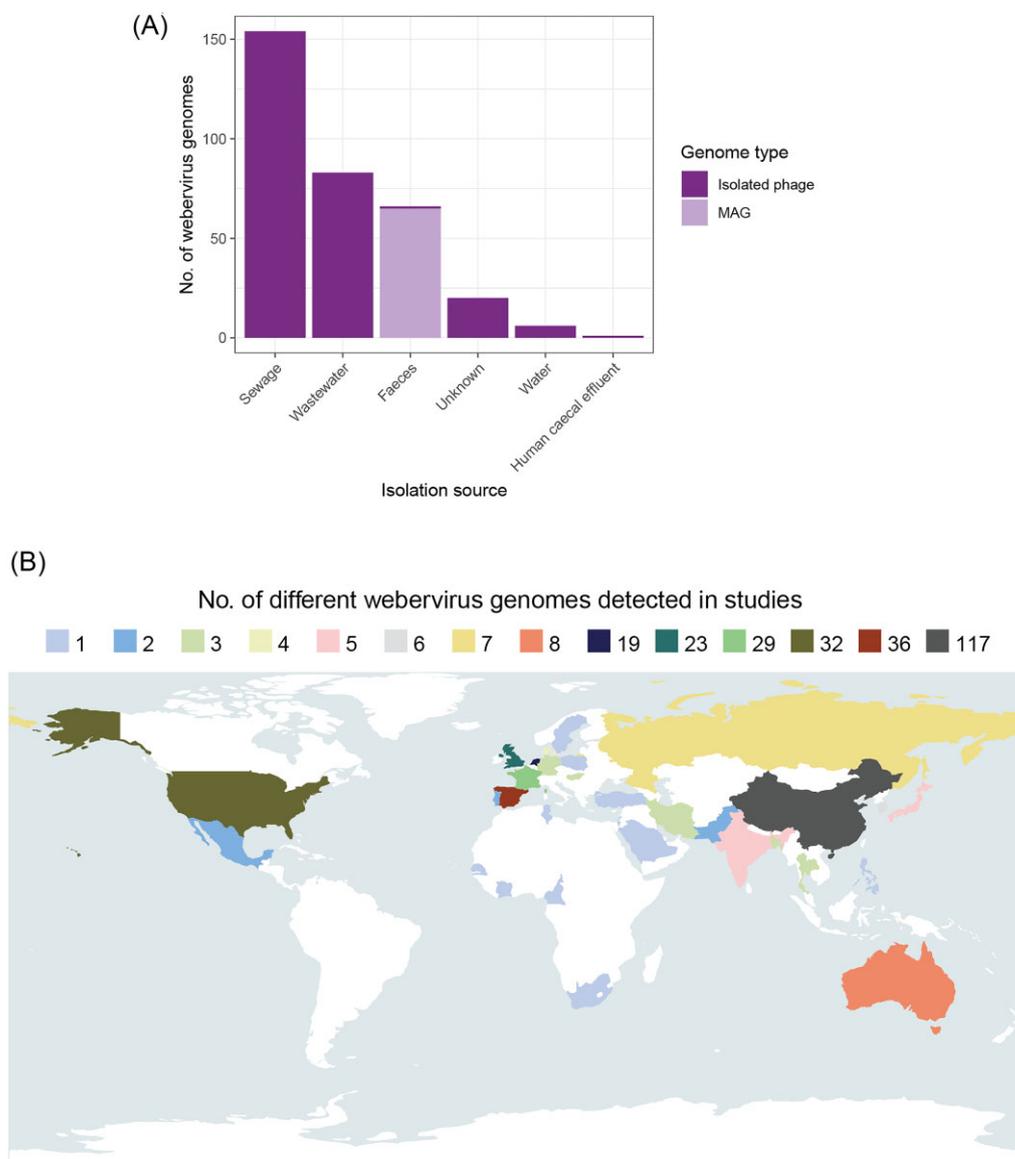


Figure 3. Distribution of weberviruses (A) Stacked bar graph showing the sources of the 330 webervirus genomes ($n = 265$ isolated phages; $n = 65$ MAGs). (B) Geographical distribution of 329 of the webervirus genomes included in this study (the location information was not available for one isolated phage, namely *Klebsiella* phage 5899STDY8049225).

Depolymerases are readily detected in webervirus genomes

As our newly isolated phages all displayed apparent depolymerase activity against one or more hosts, we aimed to identify potential depolymerases encoded within the genomes of weberviruses. Detection and characterization of these enzymes may identify standalone therapeutics or help inform on host tropism. Currently, four experimentally validated depolymerases from weberviruses have been reported in the literature: depoKP36 (Majkowska-Skrobek et al. 2016), Depo32 (Cai et al. 2023), DpK2 (Dunstan et al. 2021), and B1dep (Pertics et al. 2021). These four depolymerases were used to create a BLASTP database to interrogate the 330 webervirus genomes for similar amino acid sequences.

Using thresholds of >50% coverage, >50% identity and sequence length >800 aa, 33/330 webervirus proteomes returned hits against the validated depolymerases (Fig. 4; Supplementary Table 8). Phylogenetic analysis and amino acid identity values revealed that the depolymerases clustered into three distinct

groups, each with high bootstrap support (85%–100%; Fig. 4). Group 1 comprised four sequences and did not contain an experimentally validated depolymerase sequence. Group 2 contained four sequences, including the functionally characterized depolymerase depoKP36. Group 3 contained most of the sequences (26/33 predicted depolymerases) and included the characterized depolymerases DpK2, Depo32 and B1dep, and depolymerases encoded by four MAGs. Sequences belonging to Group 3 had a high level of conservation, as indicated by short branch lengths and sequence alignments (Supplementary Fig. E). Amino acid alignment of all 33 predicted depolymerases also revealed a high level of N-terminal sequence conservation.

Discussion

Studies from a diverse range of geographical locations have reported the isolation or detection of weberviruses from samples associated with the human gut (e.g. wastewater, sewage, fae-

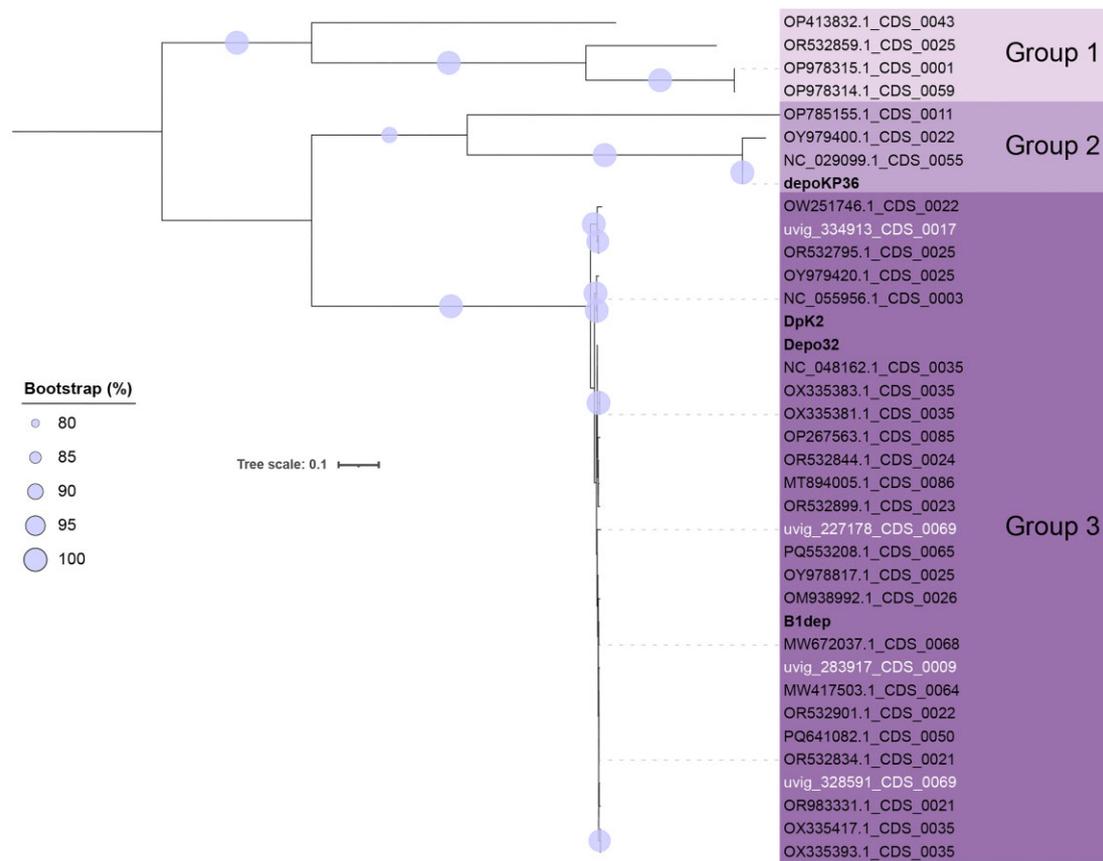


Figure 4. Phylogenetic analysis of depolymerases predicted to be encoded by weberviruses. The tree (maximum likelihood) is rooted at the midpoint. Bootstrap values are presented as a percentage of 100 replicates. Names of experimentally validated (i.e. functional) depolymerases are shown in bold black text; depolymerases predicted to be encoded by MAGs are shown in white text. Scale bar, mean number of amino acid substitutions per position.

ces, caecal effluent) (Herridge et al. 2020). To date, the majority of weberviruses have been isolated using *K. pneumoniae* as a host (Supplementary Fig. B). However, weberviruses have been reported to infect other *Klebsiella* spp., including *K. oxytoca* (Brown et al. 2017, Park et al. 2017) and *K. aerogenes* (Hudson et al. 2021). In the present study, we isolated seven new weberviruses from sewage samples, including three phages (vB_KvaS-KLPN5, vB_KvaS-KLPN6, vB_KvaS-KLPN7) that were isolated using a strain of *K. variicola* as the host (Fig. 1, Fig. 2, Supplementary Fig. A). To our knowledge, this is the first report of weberviruses infecting *K. variicola*, a recognized emerging human pathogen (Rodríguez-Medina et al. 2019) increasingly associated with carbapenem and colistin resistance (Kim et al. 2023; Li et al. 2024).

As the majority of the *Klebsiella* spp. sensitive to lysis by our weberviruses are MDR strains, the lytic phages isolated as part of this study represent attractive future therapeutics for the treatment of drug-resistant isolates belonging to the *K. pneumoniae* species complex.

In agreement with previous work (Hoyles et al. 2015, Pertics et al. 2021), the weberviruses described herein exhibited relatively narrow host ranges when screened against a panel of *Klebsiella* (including 36 clinical MDR) isolates representing a range of STs and capsule (K) types (Table 1). Phage host range is very much related to isolation host rather than phage phylogeny, with lysis appearing to be restricted based on K type. Phage-encoded depolymerases, therefore, contribute to host tropism, and previous studies have identified that weberviruses encode functionally active depolymerases (Majkowska-Skrobek et al. 2016, Dun-

stan et al. 2021, Pertics et al. 2021, Cai et al. 2023). While performing our host-range analysis, we observed the presence of haloes indicative of depolymerase activity for a small number of phage–host combinations and we, therefore, undertook a bioinformatic analysis (Fig. 4) to identify potential depolymerase enzymes encoded within webervirus genomes. Our BLASTP search identified 33 potential depolymerases which clustered into three distinct groups. The lack of an experimentally validated depolymerase sequence in Group 1 and the overall low amino acid identity shared with characterized webervirus depolymerases (<21%) make it difficult to draw conclusions related to the biological activity of these four proteins. Sequences OP978314.1_CDS_0059 and OP978315.1_CDS_0001 belong to a phage, and its evolved variant, respectively, which were characterized as part of the same study in Australia (Ngiam et al. 2024). These phages were propagated on *K. pneumoniae* 52 145 (K2: O1). According to NCBI, the isolation host of phage OP413832.1, which encodes predicted depolymerase OP413832.1_CDS_0043, is *K. pneumoniae* BS317-1 (K57: O1) (assembly accession GCF_015290145.1). No information is available for the isolation host of the phage OR532859.1, which encoded the remaining predicted Group 1 depolymerase. These data suggest that, if active, Group 1 depolymerases may hydrolyse K2 and/or K57 capsules. However, experimental validation is required.

Group 2 depolymerases are likely to be hydrolyse the K63 capsule as these sequences clustered with the experimentally validated depolymerase depoKP36, previously shown to degrade the K63 capsule of *K. pneumoniae* (Majkowska-Skrobek et al. 2016). Group 3 contained the majority of the predicted depolymerases,

and all shared high sequence similarity with the webervirus depolymerases Depo32, DpK2, and B1dep (Supplementary Table 8). These enzymes have been shown to selectively degrade the *K. pneumoniae* K2 capsule (Dunstan et al. 2021, Pertics et al. 2021, Cai et al. 2023) and are highly likely to be specific for this capsule type. The high level of sequence identity observed at the N-terminal of all the identified depolymerases is likely due to this region being responsible for anchoring the baseplate of the phage virion, and as such it is often highly conserved (Knecht et al. 2019, Latka et al. 2019). Structural analysis of Depo32 from phage GH-K3 has revealed that, in addition to the N-terminal domain, Depo32 contains a short neck helix and connection domain (residues 186–271), a β -helix domain (residues 272–642), a connection helix domain (residues 643–666), a carbohydrate-binding module (residues 667–846), and a C-terminal domain (residues 847–907) (Cai et al. 2023). It is the β -helix domain that is responsible for hydrolysis of the polysaccharide capsule. Given the high level of amino acid identity between Depo32 and the amino acid sequences comprising Group 3, it is highly likely that these potential depolymerases are structurally similar.

We were unable to identify any coding sequences in the genomes of our isolated KLPN phages sharing high similarity to the four experimentally validated webervirus depolymerase sequences used to create our BLASTP database. Thus, it is likely that any depolymerase activity associated with the phages isolated in our study is due to enzyme(s) that remain to be characterized experimentally. As part of our previous analysis of the genome of phage KLPN1, we hypothesized that ORF34 and/or ORF35 may encode the depolymerase activity of phage KLPN1, as these sequences include a predicted endo-N-acetylneuraminidase/endo-sialidase domain (Hoyles et al. 2015). Further experimental work is required to determine whether these are functionally active depolymerases. As most of the plaques we observed had no discernible haloes, it may be that alternative mechanisms are used by weberviruses for penetrating the bacterial capsule. Depolymerase-independent penetration of the capsule by *Klebsiella* phages has been reported in the literature (Beamud et al. 2023).

A ViPTree proteome-based analysis of publicly available sequence data showed 330 genomes derived from isolated phages ($n = 265$) and MAGs ($n = 65$) belonged to the genus *Webervirus*, family *Drexlerviridae* (Fig. 1). Our gene-sharing network analysis supported this finding (Fig. 2B). Taxonomic assignment of phages using whole genome gene-sharing profiles has been shown to be highly accurate; a recent study showed that vConTACT2 produces near-identical replication of existing genus-level taxonomy assignments from the ICTV (Bin Jang et al. 2019). It has been suggested that genomes comprising a genus should be evaluated by phylogenetics with the use of ‘signature genes’ that are conserved throughout all members (Turner et al. 2021). Such analyses should always produce trees that are monophyletic. Using *terL* as a ‘signature gene’, we were able to show that the genus *Webervirus* is indeed monophyletic (Fig. 2A). To assess the number of different species present within the genus, we used VIRIDIC to determine the intergenomic similarity between phage genomes (Supplementary Fig. D; Supplementary Table 4). Guidelines suggest any two phages belong to the same species if they are more than 95% identical across their entire genome (Turner et al. 2021). Genus-level separation occurs when phage genomes share <70% nucleotide identity across their genome length (Turner et al. 2021). Based on these criteria, our results show that only weberviruses belonging to Cluster 1 represent species of *Webervirus sensu stricto*. Clusters 3–9, although identified as weberviruses

using ViPTree and vConTACT2, do not represent species of *Webervirus*. The phage sequences associated with these clusters were derived from low-quality MAGs. As such, we recommend caution when using low-quality MAGs to determine taxonomic affiliations of *in silico*-generated phage sequences.

Cluster 2 phages were found to represent a novel genus (*Defiantjazzvirus*) of phage within the family *Drexlerviridae* (Figs 1 and 2, Supplementary Tables 4 and 5), with the genus *Defiantjazzvirus* most closely related to the genus *Webervirus*. All members of this novel genus reported to date infect a range of *Klebsiella* spp. (Supplementary Table 2).

As phages are among the most abundant biological entities on Earth, it is important to gain knowledge on their presence within different environments. We determined that weberviruses are distributed globally and predominated by phages associated with human faeces or water supplies contaminated with human faeces (Fig. 3). Lack of detection in most of South America and Africa is likely due to the absence of metagenomic datasets from these parts of the world rather than weberviruses not being represented in faecal samples from individuals living in countries within these regions. Compared with shotgun metagenomic datasets characterizing the total microbiota found in faeces, there are very few studies—worldwide—examining solely the intestinal virome, and PhageClouds is populated with phage genomes derived from virome datasets.

It was notable when curating our MAG dataset that none of the studies describing these data were able to predict hosts for the webervirus MAGs we have identified. Nor was the recently released tool iPhoP, specifically designed for use with MAGs (Roux et al. 2023). Our analysis using HostPhinder predicted webervirus MAGs infect *K. pneumoniae*. HostPhinder predicts the host species of a phage by searching for the most genetically similar phages in a database of reference phages with known hosts (Villarrol et al. 2016). Although the authors have shown that this whole-genome similarity-based approach is highly accurate, host range can be altered by a relatively low number of mutations, especially those localized to tail fibre proteins which are often determinants of host-cell specificity (Latka et al. 2021, Taslem Mouroso et al. 2022). In the present study, we used a strain of *K. variicola* to isolate three weberviruses, and phages of this genus have also been isolated on *K. oxytoca* and *K. aerogenes*. Although it is highly likely that 329/330 weberviruses discussed herein are phages of *Klebsiella* spp., determination of host range via plaque assays is still informative, especially when determining therapeutic utility.

Summary

We successfully characterized seven novel weberviruses that infect clinically relevant MDR *Klebsiella* spp. We have trebled the number of authenticated webervirus genomes through combining genomic data from isolated phage and MAG datasets. In doing so, we have demonstrated the importance of interrogating MAG datasets to expand the availability of curated phage genome sequences for use in genomic and ecological studies, and highlighted the need to exercise caution when assigning low-quality MAGs to taxa.

Acknowledgements

Consultant microbiologist Dr Frances Davies (Imperial College Healthcare NHS Trust) is thanked for providing access to clinical strains. We thank Emily Goren for providing mixed-liquor samples from Mogden Sewage Treatment Works (Thames Wa-

ter). Dave Baker and the QIB core sequencing team are thanked for WGS library preparation and sequencing for the PS-prefixed bacterial genomes. Horst Neve is thanked for generating TEMs of phages vB_KpnS-KLPN2, vB_KpnS-KLPN3, and vB_KpnS-KLPN4. L.H., A.L.M., and D.N. designed the study. P.S., A.L.M., S.G., T.T., and L.H. isolated and purified the phages. T.C.B. and S.J.T.D. assembled and annotated the phage genomes. A.L.M. and M.K. processed clinical isolates for whole-genome sequencing; P.S. and L.H. assembled and annotated the *Klebsiella* genomes. P.S. determined the antimicrobial profiles for the *Klebsiella* strains included in this study. F.N. determined the host ranges of the MAGs. S.J.T.D. did all bioinformatics work associated with the seven new phage genomes and the vConTACT2 analyses; L.H. did all other bioinformatics work associated with the MAGs. D.N. produced sequence data and TEM images. L.H. and A.L.M. supervised P.S. L.H. supervised T.T., T.C.B., and S.J.T.D. L.J.H. supervised M.K. D.N. supervised S.G. S.J.T.D., D.N., and L.H. drafted the manuscript. All authors read and approved the final version of the manuscript. The study of anonymized clinical isolates beyond the diagnostic requirement was approved by an NHS research ethics committee (number 06/Q0406/20).

Author contributions

Samuel J. T. Dawson (Formal analysis, Investigation, Methodology, Visualization, Writing – original draft, Writing – review & editing), Preetha Shibu (Formal analysis, Funding acquisition, Investigation, Methodology, Writing – review & editing), Sara Garnett (Data curation, Formal analysis, Investigation), Fiona Newberry (Data curation, Formal analysis, Writing – review & editing), Thomas C. Brook (Data curation, Formal analysis, Investigation, Methodology), Tobi Tijani (Data curation, Investigation, Methodology), Magdalena Kujawska (Investigation, Methodology, Writing – review & editing), Lindsay J. Hall (Funding acquisition, Resources, Supervision, Writing – review & editing), Anne L. McCartney (Conceptualization, Formal analysis, Investigation, Methodology, Resources, Supervision, Writing – review & editing), David Negus (Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Resources, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing), and Lesley Hoyles (Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing)

Supplementary data

Supplementary data are available at [FEMSEC](https://academic.oup.com/femsec) online.

Conflict of interest: None declared.

Funding

Imperial Health Charity is thanked for contributing to registration fees for the Professional Doctorate studies of P.S. P.S. was in receipt of an IBMS Research Grant (project title ‘Isolation of lytic bacteriophages active against antibiotic-resistant *Klebsiella pneumoniae*’). This work used computing resources provided by UK Med-Bio (Medical Research Council grant number MR/L01632X/1) and the Research Contingency Fund of the Department of Biosciences, Nottingham Trent University. S.J.T.D. was funded by Nottingham Trent University. S.G. completed this work as part of an MRes de-

gree at NTU. L.J.H. is supported by Wellcome Trust Investigator Awards no. [220876/Z/20/Z].

Data availability

The sequences for the seven new phage genomes described herein have been deposited in DDBJ/ENA/GenBank under accession numbers OM065837-OM065843. The genome sequences of bacteria described herein have been deposited under BioProject PRJNA917129. All supplementary material is available from https://figshare.com/projects/Weberviruses_are_gut-associated_phages_that_infect_Klebsiella_spp_/128516.

References

- Alvarez D, Merino S, Tomás JM et al. Capsular polysaccharide is a major complement resistance factor in lipopolysaccharide O side chain-deficient *Klebsiella pneumoniae* clinical isolates. *Infect Immun* 2000;**68**:953–5. <https://doi.org/10.1128/IAI.68.2.953-955.2000>
- Antimicrobial Resistance Collaborators. Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. *Lancet Lond Engl* 2022;**399**:629–55. [https://doi.org/10.1016/S0140-6736\(21\)02724-0](https://doi.org/10.1016/S0140-6736(21)02724-0)
- Bankevich A, Nurk S, Antipov D et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol J Comput Mol Cell Biol* 2012;**19**:455–77. <https://doi.org/10.1089/cmb.2012.0021>
- Beamun B, García-González N, Gómez-Ortega M et al. Genetic determinants of host tropism in *Klebsiella* phages. *Cell Rep* 2023;**42**:112048. <https://doi.org/10.1016/j.celrep.2023.112048>
- Bin Jang H, Bolduc B, Zablocki O et al. Taxonomic assignment of uncultivated prokaryotic virus genomes is enabled by gene-sharing networks. *Nat Biotechnol* 2019;**37**:632–9. <https://doi.org/10.1038/s41587-019-0100-8>
- Bolduc B, Jang HB, Doulier G et al. vConTACT: an iVirus tool to classify double-stranded DNA viruses that infect *archaea* and *bacteria*. *PeerJ* 2017;**5**:e3243. <https://doi.org/10.7717/peerj.3243>
- Bouras G, Nepal R, Houtak G et al. Pharokka: a fast scalable bacteriophage annotation tool. *Bioinforma Oxf Engl* 2023;**39**:btac776. <https://doi.org/10.1093/bioinformatics/btac776>
- Brown TL, Petrovski S, Hoyle D et al. Characterization and formulation into solid dosage forms of a novel bacteriophage lytic against *Klebsiella oxytoca*. *PLoS One* 2017;**12**:e0183510. <https://doi.org/10.1371/journal.pone.0183510>
- Cai R, Ren Z, Zhao R et al. Structural biology and functional features of phage-derived depolymerase Depo32 on *Klebsiella pneumoniae* with K2 serotype capsular polysaccharides. *Microbiol Spectr* 2023;**11**:e0530422. <https://doi.org/10.1128/spectrum.05304-22>
- Cai R, Wang Z, Wang G. et al. Biological properties and genomics analysis of vB_KpnS_GH-K3, a *Klebsiella* phage with a putative depolymerase-like protein. *Virus Genes* 2019;**55**: 696–706. <https://doi.org/10.1007/s11262-019-01681-z>
- Camarillo-Guerrero LF, Almeida A, Rangel-Pineros G et al. Massive expansion of human gut bacteriophage diversity. *Cell* 2021;**184**:1098–1109.e9. <https://doi.org/10.1016/j.cell.2021.01.029>
- Chklovski A, Parks DH, Woodcroft BJ et al. 2022. CheckM2: a rapid, scalable and accurate tool for assessing microbial genome quality using machine learning. *bioRxiv*, <https://doi.org/10.1101/2022.07.11.499243>
- Cook R, Brown N, Redgwell T et al. INfrastructure for a PHAGE Reference Database: identification of large-scale biases in the current collection of cultured phage genomes. *PHAGE* 2021;**2**:214–23. <https://doi.org/10.1089/phage.2021.0007>

- Dion MB, Plante P-L, Zufferey E et al. Streamlining CRISPR spacer-based bacterial host predictions to decipher the viral dark matter. *Nucleic Acids Res* 2021;**49**:3127–38. <https://doi.org/10.1093/nar/gkab133>
- Dunstan RA, Bamert RS, Belousoff MJ et al. Mechanistic insights into the capsule-targeting depolymerase from a *Klebsiella pneumoniae* bacteriophage. *Microbiol Spectr* 2021;**9**:e0102321. <https://doi.org/10.1128/Spectrum.01023-21>
- Fang Q, Feng Y, McNally A et al. Characterization of phage resistance and phages capable of intestinal decolonization of carbapenem-resistant *Klebsiella pneumoniae* in mice. *Commun Biol* 2022;**5**:48. <https://doi.org/10.1038/s42003-022-03001-y>
- Follador R, Heinz E, Wyres KL et al. The diversity of *Klebsiella pneumoniae* surface polysaccharides. *Microb Genomics* 2016;**2**:e000073. <https://doi.org/10.1099/mgen.0.000073>
- Gilcrease EB, Casjens SR, Bhattacharjee A et al. A *Klebsiella pneumoniae* NDM-1+ bacteriophage: adaptive polyvalence and disruption of heterogenous biofilms. *Front Microbiol* 2023;**14**:1100607. <https://doi.org/10.3389/fmicb.2023.1100607>
- Gregory AC, Zablocki O, Zayed AA et al. The Gut Virome Database reveals age-dependent patterns of virome diversity in the human gut. *Cell Host Microbe* 2020;**28**:724–740.e8. <https://doi.org/10.1016/j.chom.2020.08.003>
- Haines MEK, Hodges FE, Nale JY et al. Analysis of selection methods to develop novel phage therapy cocktails against antimicrobial resistant clinical isolates of bacteria. *Front Microbiol* 2021;**12**:613529. <https://doi.org/10.3389/fmicb.2021.613529>
- Herridge WP, Shibu P, O'Shea J et al. Bacteriophages of *Klebsiella* spp., their diversity and potential therapeutic uses. *J Med Microbiol* 2020;**69**:176–94. <https://doi.org/10.1099/jmm.0.001141>
- Hoyles L, Murphy J, Neve H et al. *Klebsiella pneumoniae* subsp. *pneumoniae*-bacteriophage combination from the caecal effluent of a healthy woman. *PeerJ* 2015;**3**:e1061. <https://doi.org/10.7717/peerj.1061>
- Hudson AT, Clark J, Gill JJ et al. Complete genome sequence of *Klebsiella aerogenes* siphophage Solomon. *Microbiol Resour Announc* 2021;**10**:e01419–20. <https://doi.org/10.1128/MRA.01419-20>
- Hyatt D, Chen G-L, Locascio PF et al. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinform* 2010;**11**:119. <https://doi.org/10.1186/1471-2105-11-119>
- Jensen TS, Opstrup KV, Christiansen G et al. Complement mediated *Klebsiella pneumoniae* capsule changes. *Microbes Infect* 2020;**22**:19–30. <https://doi.org/10.1016/j.micinf.2019.08.003>
- Kabha K, Nissimov L, Athamna A et al. Relationships among capsular structure, phagocytosis, and mouse virulence in *Klebsiella pneumoniae*. *Infect Immun* 1995;**63**:847–52. <https://doi.org/10.1128/iai.63.3.847-852.1995>
- Kim SJ, Jo J, Ko KS. Lipid A modification-induced colistin-resistant *Klebsiella variicola* from healthy adults. *J Med Microbiol* 2023;**72**. <https://doi.org/10.1099/jmm.0.001680>
- Knecht LE, Veljkovic M, Fieseler L. Diversity and function of phage encoded depolymerases. *Front Microbiol* 2019;**10**:2949. <https://doi.org/10.3389/fmicb.2019.02949>
- Lam MMC, Wick RR, Watts SC et al. A genomic surveillance framework and genotyping tool for *Klebsiella pneumoniae* and its related species complex. *Nat Commun* 2021;**12**:4188. <https://doi.org/10.1038/s41467-021-24448-3>
- Latka A, Leiman PG, Drulis-Kawa Z et al. Modeling the architecture of depolymerase-containing receptor binding proteins in *Klebsiella* phages. *Front Microbiol* 2019;**10**:2649. <https://doi.org/10.3389/fmicb.2019.02649>
- Latka A, Lemire S, Grimon D et al. Engineering the modular receptor-binding proteins of *Klebsiella* phages switches their capsule serotype specificity. *mBio* 2021;**12**:e00455–21. <https://doi.org/10.1128/mBio.00455-21>
- Lee IR, Molton JS, Wyres KL et al. Differential host susceptibility and bacterial virulence factors driving *Klebsiella* liver abscess in an ethnically diverse population. *Sci Rep* 2016;**6**:29316. <https://doi.org/10.1038/srep29316>
- Li L, Zhang Y, Wang W et al. Global emergence and genomic epidemiology of blaNDM-carrying *Klebsiella variicola*. *Infect Drug Resist* 2024;**17**:1893–901. <https://doi.org/10.2147/IDR.S460569>
- Ma Y, You X, Mai G et al. A human gut phage catalog correlates the gut phageome with type 2 diabetes. *Microbiome* 2018;**6**:24. <https://doi.org/10.1186/s40168-018-0410-y>
- Majkowska-Skrobek G, Łatka A, Berisio R et al. Capsule-targeting depolymerase, derived from *Klebsiella* KP36 phage, as a tool for the development of anti-virulent strategy. *Viruses* 2016;**8**:E324. <https://doi.org/10.3390/v8120324>
- Mangiola S, Papenfuss AT. tidyHeatmap: an R package for modular heatmap production based on tidy principles. *J Open Source Softw* 2020;**5**:2472. <https://doi.org/10.21105/joss.02472>
- Marr CM, Russo TA. Hypervirulent *Klebsiella pneumoniae*: a new public health threat. *Expert Rev Anti Infect Ther* 2019;**17**:71–3. <https://doi.org/10.1080/14787210.2019.1555470>
- Millard A, Denise R, Lestido M et al. taxmyPHAGE: automated taxonomy of dsDNA phage genomes at the genus and species level. *bioRxiv*, 2024. <https://doi.org/10.1101/2024.08.09.606593>
- Mishra CK, Choi TJ, Kang SC. Isolation and characterization of a bacteriophage F20 virulent to *Enterobacter aerogenes*. *J Gen Virol* 2012;**93**:2310–4. <https://doi.org/10.1099/vir.0.043562-0>
- Mizuta K, Ohta M, Mori M et al. Virulence for mice of *Klebsiella* strains belonging to the O1 group: relationship to their capsular (K) types. *Infect Immun* 1983;**40**:56–61. <https://doi.org/10.1128/iai.40.1.56-61.1983>
- Nayfach S, Camargo AP, Schulz F et al. CheckV assesses the quality and completeness of metagenome-assembled viral genomes. *Nat Biotechnol* 2021a;**39**:578–85. <https://doi.org/10.1038/s41587-020-00774-7>
- Nayfach S, Páez-Espino D, Call L et al. Metagenomic compendium of 189 680 DNA viruses from the human gut microbiome. *Nat Microbiol* 2021b;**6**:960–70. <https://doi.org/10.1038/s41564-021-00928-6>
- Newberry F, Shibu P, Smith-Zaitlik T et al. Lytic bacteriophage vB_KmiS-Kmi2C disrupts biofilms formed by members of the *Klebsiella oxytoca* complex, and represents a novel virus family and genus. *J Appl Microbiol* 2023;**134**:lxad079. <https://doi.org/10.1093/jambio/lxad079>
- Ngiam L, Weynberg K, Guo J. Evolutionary and co-evolutionary phage training approaches enhance bacterial suppression and delay the emergence of phage resistance. *ISME Commun* 2024;**4**:ycae082. <https://doi.org/10.1093/ismeco/ycae082>
- Nishijima S, Nagata N, Kiguchi Y et al. Extensive gut virome variation and its associations with host and environmental factors in a population-level cohort. *Nat Commun* 2022;**13**:5252. <https://doi.org/10.1038/s41467-022-32832-w>
- Nishimura Y, Yoshida T, Kuronishi M et al. ViPTree: the viral proteomic tree server. *Bioinform Oxf Engl* 2017;**33**:2379–80. <https://doi.org/10.1093/bioinformatics/btx157>
- Park E-A, Kim Y-T, Cho J-H et al. Characterization and genome analysis of novel bacteriophages infecting the opportunistic human pathogens *Klebsiella oxytoca* and *K. pneumoniae*. *Arch Virol* 2017;**162**:1129–39. <https://doi.org/10.1007/s00705-016-3202-3>
- Pertics BZ, Cox A, Nyúl A et al. Isolation and characterization of a novel lytic bacteriophage against the K2 capsule-expressing hypervirulent *Klebsiella pneumoniae* strain 52145, and identifica-

- tion of its functional depolymerase. *Microorganisms* 2021;**9**:650. <https://doi.org/10.3390/microorganisms9030650>
- Rangel-Pineros G, Millard A, Michniewski S et al. From trees to clouds: PhageClouds for fast comparison of ~640 000 phage genomic sequences and host-centric visualization using genomic network graphs. *PHAGE* 2021;**2**:194–203. <https://doi.org/10.1089/phage.2021.0008>
- Rodríguez-Medina N, Barrios-Camacho H, Duran-Bedolla J et al. *Klebsiella variicola*: an emerging pathogen in humans. *Emerg Microbes Infect* 2019;**8**:973–88. <https://doi.org/10.1080/22221751.2019.1634981>
- Roux S, Camargo AP, Coutinho FH et al. iPhoP: an integrated machine learning framework to maximize host prediction for metagenome-derived viruses of archaea and bacteria. *PLoS Biol* 2023;**21**:e3002083. <https://doi.org/10.1371/journal.pbio.3002083>
- Scholl D, Adhya S, Merrill C. *Escherichia coli* K1's capsule is a barrier to bacteriophage T7. *Appl Environ Microb* 2005;**71**:4872–4. <https://doi.org/10.1128/AEM.71.8.4872-4874.2005>
- Senhaji-Kacha A, Bernabéu-Gimeno M, Domingo-Calap P et al. Isolation and characterization of two novel bacteriophages against carbapenem-resistant *Klebsiella pneumoniae*. *Front Cell Infect Microbiol* 2024;**14**:1421724. <https://doi.org/10.3389/fcimb.2024.1421724>
- Shibu P, McCuaig F, McCartney AL et al. Improved molecular characterization of the *Klebsiella oxytoca* complex reveals the prevalence of the kleboxymycin biosynthetic gene cluster. *Microb Genomics* 2021;**7**:000592. <https://doi.org/10.1099/mgen.0.000592>
- Shibu P. *Investigations of Carbapenem-resistant Klebsiella Species and Associated Clinical Considerations*. London: Westminster, 2019.
- Smith-Zaitlik T, Shibu P, McCartney AL et al. Extended genomic analyses of the broad-host-range phages vB_KmiM-2Di and vB_KmiM-4Dii reveal slopekviruses have highly conserved genomes. *bioRxiv*, 2022. <https://doi.org/10.1101/2022.04.06.486684>
- South A. rworldmap: a new R package for mapping global data. *R J* 2011;**3**:35–43.
- Taslem Mouroso J, Awe A, Guo W et al. Understanding bacteriophage tail fiber interaction with host surface receptor: the key “blueprint” for reprogramming phage host range. *Int J Mol Sci* 2022;**23**:12146. <https://doi.org/10.3390/ijms232012146>
- Tisza MJ, Buck CB. A catalog of tens of thousands of viruses from human metagenomes reveals hidden associations with chronic diseases. *Proc Natl Acad Sci USA* 2021;**118**:e2023202118. <https://doi.org/10.1073/pnas.2023202118>
- Tsang KK, Lam MMC, Wick RR et al. Diversity, functional classification and genotyping of SHV β -lactamases in *Klebsiella pneumoniae*. *Microb Genom* 2024;**10**:001294. <https://doi.org/10.1099/mgen.0.001294>
- Turner D, Kropinski AM, Adriaenssens EM. A roadmap for genome-based phage taxonomy. *Viruses* 2021;**13**:506. <https://doi.org/10.3390/v13030506>
- Villarroel J, Kleinheinz KA, Jurtz VI et al. HostPhinder: a phage host prediction tool. *Viruses* 2016;**8**:116. <https://doi.org/10.3390/v8050116>
- Wick RR, Schultz MB, Zobel J et al. Bandage: interactive visualization of de novo genome assemblies. *Bioinforma Oxf Engl* 2015;**31**:3350–2. <https://doi.org/10.1093/bioinformatics/btv383>
- Wyres KL, Wick RR, Gorrie C et al. Identification of *Klebsiella* capsule synthesis loci from whole genome data. *Microb Genomics* 2016;**2**:e000102. <https://doi.org/10.1099/mgen.0.000102>
- Yu W-L, Ko W-C, Cheng K-C et al. Comparison of prevalence of virulence factors for *Klebsiella pneumoniae* liver abscesses between isolates with capsular K1/K2 and non-K1/K2 serotypes. *Diagn Microbiol Infect Dis* 2008;**62**:1–6. <https://doi.org/10.1016/j.diagmicrobio.2008.04.007>
- Ziller L, Blum PC, Buhl EM et al. Newly isolated *Drexlerviridae* phage LAPAZ is physically robust and fosters eradication of *Klebsiella pneumoniae* in combination with meropenem. *Virus Res* 2024;**347**:199417. <https://doi.org/10.1016/j.virusres.2024.199417>
- Zurabov F, Zhilenkov E. Characterization of four virulent *Klebsiella pneumoniae* bacteriophages, and evaluation of their potential use in complex phage preparation. *Virol J* 2021;**18**:9. <https://doi.org/10.1186/s12985-020-01485-w>