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

Samuel J. T. Dawson[1], Preetha Shibu[2]*, Sara Garnett[1], Fiona Newberry[1], Thomas C. Brook[2], Tobi Tijani[2], Magdalena Kujawska[3,4,5], Lindsay J. Hall[3,4,5], Anne L. McCartney[1], David Negus[1], Lesley Hoyles[1]

[1] Department of Biosciences, Nottingham Trent University, UK
[2] Life Sciences, University of Westminster, UK
[3] Microbes, Infection and Microbiomes, School of Infection, Inflammation and Immunology, College of Medicine and Health, University of Birmingham, UK
[4] Food, Microbiome and Health, Quadram Institute Bioscience, Norwich Research Park, Norwich, UK
[5] Norwich Medical School, University of East Anglia, Norwich Research Park, Norwich

Corresponding authors: David Negus, david.negus@ntu.ac.uk; Lesley Hoyles, lesley.hoyles@ntu.ac.uk

* **Present address:** Berkshire and Surrey Pathology Services, Frimley Health NHS Trust, Wexham Park Hospital, Slough, UK.

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

Abbreviations: MAG, metagenome-assembled genome; MDR, multidrug-resistant; ST, sequence type; TEM, transmission electron micrograph; UHGG, Unified Human Gastrointestinal Genome.

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.

ABSTRACT

Weberviruses are bacteriophages (phages) that have been shown to infect and lyse members of the genus *Klebsiella*. Due to their antimicrobial properties, phages and their products have emerged as a potential therapeutic option to help tackle increasing rates of antimicrobial resistance amongst *Klebsiella* spp. Weberviruses are an attractive option due to their high burst sizes, long shelf life and associated depolymerase enzymes. In this study we isolated and characterized seven new weberviruses and compared their genomes with those of their closest relatives to further our understanding of the genus *Webervirus*. We show that the seven phages isolated in this study can lyse a range of clinical multidrug-resistant *Klebsiella* strains, including *K. variicola*. Gene-sharing network-based analysis using vConTACT2 and incorporating all publicly available webervirus genomes (*n*=67) confirmed the seven phages as members of the genus *Webervirus* and that weberviruses are affiliated with the family *Drexlerviridae*. Using the online resource PhageClouds, we demonstrated that weberviruses can be readily detected in metagenome-derived phage datasets from a diverse range of geographical locations. Using a curated database of 67 isolated phage genomes and 60 metagenome-assembled genomes (*n*=127 total), we found that weberviruses are primarily associated with samples originating from the gut, with human faeces (61/127, 48 %), sewage (39/127, 31 %) and wastewater (14/127, 11 %) being the most frequent sources of these phages.

INTRODUCTION

Webervirus KP36 was isolated from treated sewage from a communal wastewater plant in Poland (Kęsik-Szeloch et al., 2013). This lytic phage infected ~12 % of clinical *Klebsiella pneumoniae* strains it was screened against, and exhibited depolymerase activity on its hosts (Kęsik-Szeloch et al., 2013; Majkowska-Skrobek et al., 2016). It represented the type species of a novel genus, *KP36virus*, of *Siphoviridae* belonging to the family *Tunavirinae* (Kęsik-Szeloch et al., 2013). Subsequent changes to the virus taxonomic code led to the renaming of *KP36virus* to *Webervirus* (Walker et al., 2019), with the genus *Webervirus* moved to the family *Drexliviridae* within the order *Caudovirales* in 2020 (Walker et al., 2021). Abolishment of the *Caudovirales* in July 2021 led to the family *Drexliviridae* being classified within the class *Caudoviricetes*, phylum *Uroviricota*. As of 18 January 2023, the genus *Webervirus* encompasses 32 different phage species (International Committee on Taxonomy of Viruses).

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 and long shelf life make weberviruses ideal phages to work with for biotechnological and clinical applications (Fang et al., 2022; Herridge et al., 2020). 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 target 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 lyse capsules of *Klebsiella* spp. that often hinder or make treatment with antimicrobials difficult (Cai et al., 2019; Majkowska-Skrobek et al., 2016; 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).

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). During studies to isolate additional phages that infect L4-FAA5 and multidrug-resistant (MDR) *Klebsiella* spp., we identified seven new representatives of the genus *Webervirus*. Here we describe these

new phages and compare their genomes with those of their closest relatives. We also demonstrate that 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. Genome data for clinical strains included in this study were generated as described previously (Shibu et al., 2021). Illumina and ONT Nanopore data for *K. pneumoniae* L4-FAA5 were generated by microbesNG (Birmingham, UK) as described previously (Newberry et al., 2023). *De novo* genome assembly was performed by microbesNG using Unicycler v0.4.0 (Wick et al., 2017). CheckM2 v0.1.3 (Chklovski et al., 2022) was used to confirm the quality (in terms of completeness and contamination) of all assembled genomes. PubMLST was used to assign sequence types (STs) to genomes (Jolley et al., 2018); Kaptive was used to assign capsule (K) and lipopolysaccharide (O) locus types to genomes (Wick et al., 2018).

Sample collection and processing. Sewage samples were collected from different mixed-liquor tanks at Mogden Sewage Treatment Works (Thames Water). All sewage samples were filtered-sterilized with 0.45 μm cellulose acetate filters (Millipore), and kept at 4 °C prior to use.

Isolation and propagation of phages. Filter-sterilized sewage samples were screened for phages using *Klebsiella* strains L4-FAA5, PS_misc5 and PS_misc6 (**Table 1**). Aliquots (9 ml) of each sample were inoculated into 1 ml of 10× sterile nutrient broth (Fluka) containing 50 mM $CaCl_2$ and 50 mM $MgCl_2$ and 200 μl of overnight broth culture, and incubated at 37 °C for 6–16 h. Phage-positive enrichments showed lysis (clear appearance) of the culture after incubation. Dilution series (10^{-1} to 10^{-10}) were prepared in nutrient broth from the filtered lysates of the phage-positive samples and used in double-agar overlays to obtain individual plaques as described previously (Hoyles et al., 2015).

Isolation of phage DNA

vB_KpnS-KLPN2, vB_KpnS-KLPN3 and vB_KpnS-KLPN4. Phages were precipitated from 100 ml of each lysate using PEG 8000/6 % NaCl and DNA isolated from the three phages as described previously (Hoyles et al., 2015). Integrity and quality of DNA were checked using agarose gel electrophoresis and Nanodrop.

vB_KvaS-KLPN5, vB_KvaS-KLPN6, vB_KvaS-KLPN7 and vB_KpnS-KLPN8. These phages 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 as described previously (Smith-Zaitlik et al., 2022), with images analysed using ImageJ (Schindelin et al., 2012).

Phage genome sequencing, assembly and annotation. Illumina sequence data for phages vB_KpnS-KLPN2, vB_KpnS-KLPN3 and vB_KpnS-KLPN4 were generated by microbesNG in 2016. Data quality for the provided trimmed sequence data were checked using FastQC v0.11.9 and did not need further trimming. 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 according to (Smith-Zaitlik et al., 2022) (Illumina MiSeq Reagent Kit v2; 150-bp paired-end reads). 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 v0.8.1 (checkv-db-v1.0; (Nayfach et al., 2021a)) was used to determine contamination and completeness of the genomes.

Genes in all phage genomes included in this study (**Table 2**) were predicted using the PHROG (Terzian et al., 2021) hmm dataset and Prokka v1.14.6 (Seemann, 2014).

Comparison of webervirus genomes. ViPTree v1.9.1 (Nishimura et al., 2017) was used to determine whether the seven phage genomes were closely related to previously described double-stranded DNA viruses. Initial analyses showed them to be closely related to *Webervirus KP36* (Kęsik-Szeloch et al., 2013) and other weberviruses. As the online ViPTree database does not contain genomes of all members of the genus *Webervirus*, we curated a list of all known webervirus sequences within NCBI GenBank in April 2022 (**Table**

2). The sequence files were concatenated and uploaded to VIRIDIC web (http://rhea.icbm.uni-oldenburg.de/VIRIDIC/) (Moraru et al., 2020). They were also added to the vConTACT v2.0 database (Bin Jang et al., 2019; Bolduc et al., 2017). The similarity matrix created using VIRIDIC was visualized using heatmap.2 in gplots v3.1.1. The gene-sharing network created using vConTACT v2.0 was imported into Gephi v0.9.2 (Bastian et al., 2009) for visualization.

Large-subunit terminase genes were extracted from genomes using Biostrings v2.62.0 (Pagès et al., 2021). A multiple-sequence alignment was created for the genes (Clustal Omega 1.2.2 implemented in Geneious Prime v2020.0.5; options – group sequences by similarity, 5 representative iterations). This alignment was used to create a bootstrapped maximum-likelihood tree. Phangorn v2.7.1 and Ape v5.5 (Paradis and Schliep, 2019; Schliep et al., 2017) were used to generate the trees using the LG model (Le and Gascuel, 2008).

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. 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 (Camarillo-Guerrero et al., 2021; Gregory et al., 2020; Tisza and Buck, 2021)) were recovered from the relevant datasets. CheckV v0.8.1 (checkv-db-v1.0; (Nayfach et al., 2021a)) was used to determine contamination and completeness of our curated dataset; those genomes that were complete and not already included in checkv-db-v1.0 were added to the database (**Supplementary Table 2**), which was then used to determine the quality (completeness, contamination) of the MAGs identified as being closely related to *Webervirus KLPN1*.

The MAG sequences were analysed using ViPTree v2.0 (online) to confirm their affiliation with the genus *Webervirus*. They were also included in a vConTACT analysis with our curated set of 67 webervirus genomes.

Distribution of weberviruses. The global distribution of weberviruses was determined by identifying the source and location information for the 67 GenBank genomes (**Table 2**) and the 60 MAGs (**Supplementary Table 3**). Data were aggregated based on isolation source or

geographical location, with these latter data visualized using the R package rworldmap v1.3.6 (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. The 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 1,846,441 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 ≥95 % of the spacer length.

Identification of potential depolymerases among weberviruses. A BLASTP database was created using amino acid sequences from experimentally validated *Webervirus* depolymerases and a BLASTP search was ran versus all webervirus genomes. Sequences used to build the BLAST database can be found in **Supplementary Table 4**. MUSCLE alignments were created in Geneious prime (v2023.1.2).

RESULTS

Initial characterization of the seven new phages

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). Strain L4-FAA5 (K2:O1v1, ST380) was originally isolated from human caecal effluent along with *Webervirus KLPN1* (Hoyles et al., 2015), while strains PS_misc5 (K81:OL101; gapA-16, infB-24, mdh-21, pgi-31, phoE-47, rpoB-22, tonB-281) and PS_misc6 (K110:O2v1, 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 siphovirus-type morphology (**Supplementary Figure A**). Host-range analysis showed the seven phages had different infection profiles (**Table 1**).

Genome-based analyses of the seven new phages

Whole-genome sequences were generated for the seven phages. Bandage (data not shown) and CheckV (**Supplementary Table 2**) analyses confirmed all the genomes were circular and complete. None of the genomes was contaminated. An initial online ViPTree analysis showed the new genomes belonged to the genus *Webervirus* (data not shown). All publicly available webervirus genomes (available as of April 2022) were downloaded from GenBank to allow comparison with our newly sequenced phages.

A gene-sharing network was created with 67 webervirus genomes derived from isolated phages (**Table 2**) using vConTACT v2.0 (Bin Jang et al., 2019; Bolduc et al., 2017). The network had 2,619 nodes and 51,589 edges. It was filtered based on the giant component topology, removing all nodes that were not connected to the main cluster (**Figure 1**). This left 1,523 nodes (58.15 % of 2,619) and 33,573 edges (65.08 % of 51,589) visible. The vConTACT-based analysis confirmed the affiliation of the seven phages with the family *Drexlerviridae* (**Figure 1**). The monophyletic nature of the genus *Webervirus* – including the seven new phages – was confirmed by analysis of large-subunit terminase genes (**Figure 2**).

VIRIDIC (**Supplementary Figure B**) data suggested there were three species within the genus *Webervirus*, based on the clustering of the 67 genomes.

Detection of weberviruses in metagenomic 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. Of the 67 webervirus genomes in our curated database, 47 were included in the PhageClouds output. Five PhageClouds hits with the prefix PIGEON (Ter Horst et al., 2021) represented phages already included in our curated dataset (**Supplementary Table 5**), so were excluded from further analyses. 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; an unsurprising finding as both MAGs are derived from the same dataset (Ma et al., 2018)). 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

retained for further analyses. This left us with 60 MAGs (available as 60_MAG_sequences.fasta in **Supplementary Material**), potentially representing weberviruses, that were subject to further analyses.

The MAGs ranged from 10,230 to 55,276 nt (mean 41,737 nt) in length (**Supplementary Table 6**). Thirteen of our 67 curated genomes were determined to be complete (CheckV) and of high quality (MIUVIG) by CheckV so were added to the CheckV database (**Supplementary Table 2**). The updated database was then used to assess the quality (completeness, contamination) of the 60 MAGs. Twenty-seven MAGs were complete/high quality, 15 were high quality/high quality, eight were medium quality/genome fragments and 10 were low quality/genome fragments (**Supplementary Table 6**).

The 60 MAGs were uploaded to ViPTree v2.0 for analysis. In a comparison of 2,981 sequences, all MAGs clustered with the genus *Webervirus* (**Supplementary Material**). vConTACT analysis confirmed the association of the MAGs with the genus *Webervirus*, with only weberviruses represented in the corresponding module (**Figure 3a**).

Distribution and diversity of weberviruses

Our dataset represented a total of 67 isolated phages and 60 MAGs. Weberviruses were recovered from a variety of sample types, with human faeces (61/127, 48 %), sewage (39/127, 31 %) and wastewater (14/127, 11 %) the most frequent sources of these phages (**Figure 3b**, **Supplementary Table 7**). Seven (5.5 %) of the weberviruses were of unknown origin. Only one genome was recovered from each of the other sample types (activated sludge, human caecal effluent, different water types).

Of the 127 different webervirus genomes included in this study (*n*=67 isolated phage, *n*=60 MAGs), 126 were of known geographical origin (**Figure 3c**). According to our analysis, weberviruses have been detected most frequently in samples emanating from: China (*n*=54); USA (*n*=24); UK (*n*=10); Spain (*n*=8); Fiji (*n*=5); Denmark and Russia (*n*=4 each); South Korea (*n*=3); India, Hungary and Austria (*n*=2 each); Turkey, Thailand, Sweden, Singapore, Poland, Pakistan, Cote d'Ivoire and Cameroon (*n*=1 each). The majority (108/127, 85 %) of webervirus genomes appear to have been detected only once. However, for MAGs described by the GPD (uvig prefix) (Camarillo-Guerrero et al., 2021), data were available with respect to the number of metagenomic samples each had been recovered from: 11 had been recovered from two samples each, one had been recovered from four samples, two had been recovered from eight samples each, and five had been recovered from 16 samples each (**Supplementary Table 3**). Recovery of webervirus genomes from one more than

sample was country-specific (i.e. for the five genomes recovered from 16 samples, all derived from Fiji).

ViPTree analysis of all 127 genomes showed clustering of several phage genomes based on geographical origin (**Supplementary Figure C**).

Phage–host prediction for the MAGs

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*). HostPhinder 1.1 (Villarroel 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 60 *Webervirus* MAGs (**Supplementary Table 8**). All were predicted to infect *Klebsiella pneumoniae*.

Depolymerases of weberviruses

As our newly isolates phages all displayed apparent depolymerase activity against one or more hosts, we aimed to identify potential depolymerases encoded within the genomes of 127 high-quality and complete webervirus genomes. 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 127 phage genomes for similar amino acid sequences.

In addition to the four previously reported sequences, our search identified four additional potential depolymerases encoded within separate webervirus genomes, with two of these belonging to metagenome-assembled genomes (**Supplementary Table 9**). All four sequences shared >98 % amino acid identity and 100 % sequence coverage with the K2 capsule-degrading depolymerases Depo32, DpK2 and B1dep, but less than 64 % amino acid over < 36 % query coverage with depolymerase depoKP36, an enzyme specific for the K63 capsular type. An amino acid similarity matrix displaying percent amino acid identity between all eight full-length sequences in included in (**Supplementary Table 10**). Alignment of the depolymerase amino acid sequences highlighted a high level of N-terminal conservation (**Supplementary Figure D**).

DISCUSSION

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. The difficulty of treating such isolates with conventional antibiotics has resulted in the investigation of new therapeutic modalities, including phages and their products. We have previously published an extensive review on *Klebsiella* phage and their potential as therapeutics (Herridge et al., 2020). To realize the potential of phage therapy, it is important to comprehensively characterize phages with clinical utility. Here, we describe the isolation and characterization of seven new phages belonging to the genus *Webervirus* with lytic activity against clinical MDR *Klebsiella* spp. and characterize these in the context of other available *Webervirus* genome sequences.

Isolation and characterization of seven new weberviruses

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, faeces, caecal effluent) (Herridge et al., 2020). To date, the majority of weberviruses have been isolated using *K. pneumoniae* as a host. 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. 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). *K. variicola* PS_misc5 (this study) is a carbapenem-resistant clinical isolate that encodes the class D β-lactamase OXA-48. The remaining four phages were isolated on either the caecal isolate *K. pneumoniae* L4-FAA5 (vB_KpnS-KLPN2, vB_KpnS-KLPN3, vB_KpnS-KLPN4) or the clinical isolate *K. pneumoniae* PS_misc6 (vB_KpnS-KLPN8). Strain PS_misc6 encodes the carbapenem-degrading metallo-β-lactamase NDM, while L4-FAA5 is resistant to cefotaxime, temocillin, ceftoxitin and aztreonam. The lytic phages isolated as part of this study may, therefore, 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**, **Supplementary 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. Phages *Webervirus KLPN1*, vB_KpnS-KLPN2, vB_KpnS-KLPN3 and vB_KpnS-KLPN4 had the broadest host ranges and were all isolated on the same host – *K. pneumoniae* L4-FAA5 (K2). Based on our phylogenetic analysis of phage-encoded large-subunit terminase amino acid sequences *Webervirus KLPN1* belongs to cluster 1 whereas vB_KpnS-KLPN2, vB_KpnS-KLPN3 and vB_KpnS-KLPN4 belong to cluster 2 (**Figure 2**). Phages *Webervirus KLPN1*, KpnS-KLPN2 and vB_KpnS-KLPN3 shared identical host ranges, showing lytic activity against the same 15 strains from the collection of 37 isolates (phage vB_KpnS-KLPN4 exhibited a near identical host range but lysed one strain fewer) (**Table 1**). All four phages were restricted to hosts predicted to produce either the K2 (14/15 strains lysed) or K64 (1/15 strains lysed) capsule. Phages vB_KvaS-KLPN5, vB_KvaS-KLPN6 and vB_KvaS-KLPN7 were isolated on *K. variicola* PS_misc5 (K81) and shared identical, narrow host ranges; infecting only a single isolate (PS_Kpn24; K2) other than their original host. Phages vB_KvaS-KLPN5 and vB_KvaS-KLPN7 were identified as belonging to cluster 1, whereas vB_KvaS-KLPN6 groups with cluster 2. Phage vB_KpnS-KLPN8 was restricted to lysis of its original host strain, *K. pneumoniae* PS_misc6 and is a member of cluster 2 (**Figure 2**).

Capsular (K) type has 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 (Dunstan et al., 2021; Scholl et al., 2005). To overcome this physical barrier, phages encode enzymes – frequently referred to as depolymerases – that selectively degrade polysaccharides or polypeptides that comprise the bacterial capsule (Cai et al., 2023; Dunstan et al., 2021; Hoyles et al., 2015; Majkowska-Skrobek et al., 2016; Pertics et al., 2021). Phage-encoded depolymerases, therefore, contribute to host tropism and previous studies have identified that weberviruses encode functionally active depolymerase enzymes (Cai et al., 2023; Dunstan et al., 2021; Majkowska-Skrobek et al., 2016; Pertics et al., 2021). While performing our host-range analysis, we observed the presence of haloes indicative of depolymerase activity for a small number of phage–host combinations (haloes were was associated with 18 % of the plaques; **Table 1**). We, therefore, undertook a bioinformatic analysis to identify potential depolymerase enzymes encoded within webervirus genomes. Our BLASTP search identified four potential depolymerases that shared significant sequence similarity with the webervirus depolymerases Depo32, DpK2 and B1dep. These enzymes have been shown to selectively

degrade the *K. pneumoniae* K2 capsule (Cai et al., 2023; Dunstan et al., 2021; Pertics et al., 2021). As the four sequences shared far less similarity with the third depolymerase included in our BLAST database, a webervirus enzyme which hydrolyses the K63 capsule, the identified depolymerases may have specificity towards the K2 capsule. All depolymerases shared a high level of sequence identity at the N-terminal region (**Supplementary Figure D**). The N-terminal region of phage-associated depolymerases is typically responsible for anchoring the baseplate of the phage virion and 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 four additional sequences identified, 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 three 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/endosialidase 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).

Genome-based analyses of weberviruses

Our gene-sharing network analysis of 67 webervirus genomes derived from isolated phages confirmed their association with the family *Drexlerviridae* (**Figure 1**). 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 viral taxonomy assignments from the International Committee on Taxonomy of Viruses (ICTV) (Bin Jang et al., 2019). It has been suggested that genomes comprising a

genus should be evaluated by phylogenetics with the use of conserved 'signature genes' that are conserved throughout all members (Turner et al., 2021). Such analyses should always produce trees that are monophyletic. Using the large-subunit terminase as a 'signature gene', we were able to show that the genus *Webervirus* is indeed monophyletic (**Figure 2**). To assess the number of different species present within the genus, we used VIRIDIC to determine the intergenomic similarity between phage genomes (**Supplementary Figure B**). 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). Based on these criteria, our results show that weberviruses cluster into three distinct species.

As phages are among the most abundant biological entities on Earth, it is important to gain knowledge on their presence within different environments. To better understand the wider diversity and distribution of weberviruses, we searched the genome sequence of *Webervirus KLPN1* against metagenomic datasets using the online resource PhageClouds (Rangel-Pineros et al., 2021). We used PhageClouds previously to demonstrate the presence of slopekviruses in human faeces (Smith-Zaitlik et al., 2022). We identified 60 high-quality webervirus MAGs in a range of samples, thereby doubling the number of webervirus genomes available to the research community for genomic analyses (**Supplementary Table 3**). Their affiliation with the genus *Webervirus* was confirmed by our vConTACT2 (**Figure 3a**) and phylogenetic analyses (**Supplementary Figure C**). We also determined that weberviruses are distributed globally and predominated by phages associated with human faeces or water supplies contaminated with human faeces (**Figure 3**). Lack of detection in South America is likely due to absence of metagenomic datasets from this part of the world rather than weberviruses not being represented in faecal samples from individuals living in countries within this region. 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 was able to predict hosts for the webervirus MAGs we have identified. Our analysis using HostPhinder predicts that all 60 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 (Villarroel 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 Mourosi 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 all 127 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 *Klebsiella* spp. We have doubled the number of available 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.

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 Water). Dave Baker and the QIB core sequencing team are thanked for WGS library preparation and sequencing for the PS-prefixed bacterial genomes. Andrew Millard (University of Leicester) is thanked for providing us with the PHROG hmm dataset for use with Prokka. Horst Neve is thanked for generating TEMs of phages vB_KpnS-KLPN2, vB_KpnS-KLPN3 and vB_KpnS-KLPN4.

LH, ALM and DN designed the study. PS, ALM, SG, TT and LH isolated and purified the phages. TCB and SJTD assembled and annotated the phage genomes. ALM and MK processed clinical isolates for whole-genome sequencing; PS and LH assembled and annotated the *Klebsiella* genomes. PS determined the antimicrobial profiles for the *Klebsiella* strains included in this study. FN determined the host ranges of the MAGs. SJTD did all bioinformatics work associated with the seven new phage genomes and the vConTACT analyses; LH did all other bioinformatics work associated with the MAGs. DN produced sequence data and TEM images. LH and ALM supervised PS. LH supervised TT, TCB and SJTD. DN supervised SG. SJTD, DN and LH drafted the manuscript. All authors read and approved the final version of the manuscript.

FUNDING

Imperial Health Charity is thanked for contributing to registration fees for the Professional Doctorate studies of PS. PS 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. SJTD was funded by Nottingham Trent University. SG completed this work as part of an MRes degree at NTU. LJH is funded by Wellcome Trust Investigator Award no. 220876/Z/20/Z.

CONFLICTS OF INTEREST

The authors declare that there are no conflicts of interest.

ETHICS

The study of anonymised clinical isolates beyond the diagnostic requirement was approved by an NHS research ethics committee (number 06/Q0406/20).

REFERENCES

Antimicrobial Resistance Collaborators, 2022. Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. Lancet Lond. Engl. 399, 629–655. https://doi.org/10.1016/S0140-6736(21)02724-0

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., Pyshkin, A.V., Sirotkin, A.V., Vyahhi, N., Tesler, G., Alekseyev, M.A., Pevzner, P.A., 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J. Comput. Biol. J. Comput. Mol. Cell Biol. 19, 455–477. https://doi.org/10.1089/cmb.2012.0021

Bastian, M., Heymann, S., Jacomy, M., 2009. Gephi: an open source software for exploring and manipulating networks, in: International AAAI Conference on Weblogs and Social Media.

Beamud, B., García-González, N., Gómez-Ortega, M., González-Candelas, F., Domingo-Calap, P., Sanjuan, R., 2023. Genetic determinants of host tropism in Klebsiella phages. Cell Rep. 42, 112048. https://doi.org/10.1016/j.celrep.2023.112048

Bin Jang, H., Bolduc, B., Zablocki, O., Kuhn, J.H., Roux, S., Adriaenssens, E.M., Brister, J.R., Kropinski, A.M., Krupovic, M., Lavigne, R., Turner, D., Sullivan, M.B., 2019. Taxonomic assignment of uncultivated prokaryotic virus genomes is enabled by gene-sharing networks. Nat. Biotechnol. 37, 632–639. https://doi.org/10.1038/s41587-019-0100-8

Bleriot, I., Blasco, L., Pacios, O., Fernández-García, L., Ambroa, A., López, M., Ortiz-Cartagena, C., Cuenca, F.F., Oteo, J., Pascual, Á., Martínez-Martínez, L., Domingo-Calap, P., Wood, T.K., Tomás, M., 2021. «PemIK (PemK/PemI) type II TA system from

Klebsiella pneumoniae clinical strains inhibits lytic phage». https://doi.org/10.21203/rs.3.rs-679460/v1

Bolduc, B., Jang, H.B., Doulcier, G., You, Z.-Q., Roux, S., Sullivan, M.B., 2017. vConTACT: an iVirus tool to classify double-stranded DNA viruses that infect Archaea and Bacteria. PeerJ 5, e3243. https://doi.org/10.7717/peerj.3243

Brown, T.L., Petrovski, S., Hoyle, D., Chan, H.T., Lock, P., Tucci, J., 2017. Characterization and formulation into solid dosage forms of a novel bacteriophage lytic against Klebsiella oxytoca. PloS One 12, e0183510. https://doi.org/10.1371/journal.pone.0183510

Cai, R., Ren, Z., Zhao, R., Lu, Y., Wang, X., Guo, Z., Song, J., Xiang, W., Du, R., Zhang, X., Han, W., Ru, H., Gu, J., 2023. Structural biology and functional features of phage-derived depolymerase Depo32 on Klebsiella pneumoniae with K2 serotype capsular polysaccharides. Microbiol. Spectr. 11, e0530422. https://doi.org/10.1128/spectrum.05304-22

Cai, R., Wang, Z., Wang, G., Zhang, H., Cheng, M., Guo, Z., Ji, Y., Xi, H., Wang, X., Xue, Y., Ur Rahman, S., Sun, C., Feng, X., Lei, L., Tong, Y., Han, W., Gu, J., 2019. Biological properties and genomics analysis of vB_KpnS_GH-K3, a Klebsiella phage with a putative depolymerase-like protein. Virus Genes. https://doi.org/10.1007/s11262-019-01681-z

Camarillo-Guerrero, L.F., Almeida, A., Rangel-Pineros, G., Finn, R.D., Lawley, T.D., 2021. Massive expansion of human gut bacteriophage diversity. Cell 184, 1098-1109.e9. https://doi.org/10.1016/j.cell.2021.01.029

Cao, F., Wang, X., Wang, Linhui, Li, Z., Che, J., Wang, Lili, Li, X., Cao, Z., Zhang, J., Jin, L., Xu, Y., 2015. Evaluation of the efficacy of a bacteriophage in the treatment of pneumonia induced by multidrug resistance Klebsiella pneumoniae in mice. BioMed Res. Int. 2015, 752930. https://doi.org/10.1155/2015/752930

Castillo, M., Tran, R., Newkirk, H., Liu, M., Gill, J.J., Ramsey, J., 2019. Complete Genome Sequence of Sin4, a Siphophage Infecting Carbapenemase-Producing Klebsiella pneumoniae. Microbiol. Resour. Announc. 8. https://doi.org/10.1128/MRA.01048-19

Chklovski, A., Parks, D.H., Woodcroft, B.J., Tyson, G.W., 2022. CheckM2: a rapid, scalable and accurate tool for assessing microbial genome quality using machine learning. https://doi.org/10.1101/2022.07.11.499243

Cranston, A., Danielson, P., Arens, D.K., Barker, A., Birch, E.K., Brown, H., Carr, E., Cero, P., Chow, J., Correa, E., Dean, J., Dunn, M., Eberhard, N., Egbert, A., Foster, K., Gaertner, R., Gleave, A., Gomez, A., Gordon, J.B., Harris, E.B., Heaps, C., Hyer, M., Johnson, A., Johnson, L., Kim, M., Kruger, J.L., Leonard, T., LeSueur, A., Lima, S., Marshall, N., Moulton, R., Newey, C.R., Owen, D., Packard, A., Rolfson, A., Suorsa, A.R., Rodriguez, W., Sandoval, C., Sharma, R., Smith, A., Sork, C., Soule, C., Soule, S., Stewart, J., Stoker, T., Thompson, D.W., Thurgood, T., Walker, J., Zaugg, E., Casjens, S.R., Grose, J.H., 2022. Genome Sequences of 22 T1-like Bacteriophages That Infect Enterobacteriales. Microbiol. Resour. Announc. 11, e0122121. https://doi.org/10.1128/mra.01221-21

Dion, M.B., Plante, P.-L., Zufferey, E., Shah, S.A., Corbeil, J., Moineau, S., 2021. Streamlining CRISPR spacer-based bacterial host predictions to decipher the viral dark matter. Nucleic Acids Res. 49, 3127–3138. https://doi.org/10.1093/nar/gkab133

Dunstan, R.A., Bamert, R.S., Belousoff, M.J., Short, F.L., Barlow, C.K., Pickard, D.J., Wilksch, J.J., Schittenhelm, R.B., Strugnell, R.A., Dougan, G., Lithgow, T., 2021. Mechanistic Insights into the Capsule-Targeting Depolymerase from a Klebsiella pneumoniae

Bacteriophage. Microbiol. Spectr. 9, e0102321. https://doi.org/10.1128/Spectrum.01023-21

Fang, Q., Feng, Y., McNally, A., Zong, Z., 2022. Characterization of phage resistance and phages capable of intestinal decolonization of carbapenem-resistant Klebsiella pneumoniae in mice. Commun. Biol. 5, 48. https://doi.org/10.1038/s42003-022-03001-y

Gao, S., Linden, S.B., Nelson, D.C., 2017. Complete Genome Sequence of Klebsiella pneumoniae Phages SopranoGao, MezzoGao, and AltoGao. Genome Announc. 5. https://doi.org/10.1128/genomeA.01009-17

Gramer, J., Kenny, S., Newkirk, H., Liu, M., Gill, J.J., Ramsey, J., 2019. Complete Genome Sequence of Klebsiella pneumoniae Siphophage Skenny. Microbiol. Resour. Announc. 8. https://doi.org/10.1128/MRA.01036-19

Gregory, A.C., Zablocki, O., Zayed, A.A., Howell, A., Bolduc, B., Sullivan, M.B., 2020. The Gut Virome Database Reveals Age-Dependent Patterns of Virome Diversity in the Human Gut. Cell Host Microbe 28, 724-740.e8. https://doi.org/10.1016/j.chom.2020.08.003

Haines, M.E.K., Hodges, F.E., Nale, J.Y., Mahony, J., van Sinderen, D., Kaczorowska, J., Alrashid, B., Akter, M., Brown, N., Sauvageau, D., Sicheritz-Pontén, T., Thanki, A.M., Millard, A.D., Galyov, E.E., Clokie, M.R.J., 2021. Analysis of Selection Methods to Develop Novel Phage Therapy Cocktails Against Antimicrobial Resistant Clinical Isolates of Bacteria. Front. Microbiol. 12, 613529. https://doi.org/10.3389/fmicb.2021.613529

Herridge, W.P., Shibu, P., O'Shea, J., Brook, T.C., Hoyles, L., 2020. Bacteriophages of Klebsiella spp., their diversity and potential therapeutic uses. J. Med. Microbiol. 69, 176–194. https://doi.org/10.1099/jmm.0.001141

Hesse, S., Rajaure, M., Wall, E., Johnson, J., Bliskovsky, V., Gottesman, S., Adhya, S., 2020. Phage Resistance in Multidrug-Resistant Klebsiella pneumoniae ST258 Evolves via Diverse Mutations That Culminate in Impaired Adsorption. mBio 11, e02530-19. https://doi.org/10.1128/mBio.02530-19

Horváth, M., Kovács, T., Koderivalappil, S., Ábrahám, H., Rákhely, G., Schneider, G., 2020. Identification of a newly isolated lytic bacteriophage against K24 capsular type, carbapenem resistant Klebsiella pneumoniae isolates. Sci. Rep. 10, 5891. https://doi.org/10.1038/s41598-020-62691-8

Hoyles, L., Murphy, J., Neve, H., Heller, K.J., Turton, J.F., Mahony, J., Sanderson, J.D., Hudspith, B., Gibson, G.R., McCartney, A.L., van Sinderen, D., 2015. Klebsiella pneumoniae subsp. pneumoniae-bacteriophage combination from the caecal effluent of a healthy woman. PeerJ 3, e1061. https://doi.org/10.7717/peerj.1061

Hudson, A.T., Clark, J., Gill, J.J., Liu, M., 2021. Complete Genome Sequence of Klebsiella aerogenes Siphophage Solomon. Microbiol. Resour. Announc. 10, e01419-20. https://doi.org/10.1128/MRA.01419-20

Jolley, K.A., Bray, J.E., Maiden, M.C.J., 2018. Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications. Wellcome Open Res. 3, 124. https://doi.org/10.12688/wellcomeopenres.14826.1

Kęsik-Szeloch, A., Drulis-Kawa, Z., Weber-Dąbrowska, B., Kassner, J., Majkowska-Skrobek, G., Augustyniak, D., Lusiak-Szelachowska, M., Zaczek, M., Górski, A., Kropinski, A.M., 2013. Characterising the biology of novel lytic bacteriophages infecting multidrug resistant Klebsiella pneumoniae. Virol. J. 10, 100. https://doi.org/10.1186/1743-422X-10-100

Kim, S.J., Jo, J., Ko, K.S., 2023. Lipid A modification-induced colistin-resistant Klebsiella variicola from healthy adults. J. Med. Microbiol. 72. https://doi.org/10.1099/jmm.0.001680

Knecht, L.E., Veljkovic, M., Fieseler, L., 2019. Diversity and Function of Phage Encoded Depolymerases. Front. Microbiol. 10, 2949. https://doi.org/10.3389/fmicb.2019.02949

Latka, A., Leiman, P.G., Drulis-Kawa, Z., Briers, Y., 2019. Modeling the Architecture of Depolymerase-Containing Receptor Binding Proteins in Klebsiella Phages. Front. Microbiol. 10, 2649. https://doi.org/10.3389/fmicb.2019.02649

Latka, A., Lemire, S., Grimon, D., Dams, D., Maciejewska, B., Lu, T., Drulis-Kawa, Z., Briers, Y., 2021. Engineering the Modular Receptor-Binding Proteins of Klebsiella Phages Switches Their Capsule Serotype Specificity. mBio 12, e00455-21. https://doi.org/10.1128/mBio.00455-21

Le, S.Q., Gascuel, O., 2008. An improved general amino acid replacement matrix. Mol. Biol. Evol. 25, 1307–1320. https://doi.org/10.1093/molbev/msn067

Li, L., Zhang, Y., Wang, W., Chen, Y., He, F., Yu, Y., 2024. Global Emergence and Genomic Epidemiology of blaNDM-Carrying Klebsiella variicola. Infect. Drug Resist. 17, 1893–1901. https://doi.org/10.2147/IDR.S460569

Ma, Y., You, X., Mai, G., Tokuyasu, T., Liu, C., 2018. A human gut phage catalog correlates the gut phageome with type 2 diabetes. Microbiome 6, 24. https://doi.org/10.1186/s40168-018-0410-y

Majkowska-Skrobek, G., Łątka, A., Berisio, R., Maciejewska, B., Squeglia, F., Romano, M., Lavigne, R., Struve, C., Drulis-Kawa, Z., 2016. Capsule-Targeting Depolymerase, Derived from Klebsiella KP36 Phage, as a Tool for the Development of Anti-Virulent Strategy. Viruses 8, E324. https://doi.org/10.3390/v8120324

Martinez, N., Williams, E., Newkirk, H., Liu, M., Gill, J.J., Ramsey, J., 2019. Complete Genome Sequence of Klebsiella pneumoniae Phage Sweeny. Microbiol. Resour. Announc. 8. https://doi.org/10.1128/MRA.01047-19

Mishra, C.K., Choi, T.J., Kang, S.C., 2012. Isolation and characterization of a bacteriophage F20 virulent to Enterobacter aerogenes. J. Gen. Virol. 93, 2310–2314. https://doi.org/10.1099/vir.0.043562-0

Moraru, C., Varsani, A., Kropinski, A.M., 2020. VIRIDIC—A Novel Tool to Calculate the Intergenomic Similarities of Prokaryote-Infecting Viruses. Viruses 12, 1268. https://doi.org/10.3390/v12111268

Nayfach, S., Camargo, A.P., Schulz, F., Eloe-Fadrosh, E., Roux, S., Kyrpides, N.C., 2021a. CheckV assesses the quality and completeness of metagenome-assembled viral genomes. Nat. Biotechnol. 39, 578–585. https://doi.org/10.1038/s41587-020-00774-7

Nayfach, S., Páez-Espino, D., Call, L., Low, S.J., Sberro, H., Ivanova, N.N., Proal, A.D., Fischbach, M.A., Bhatt, A.S., Hugenholtz, P., Kyrpides, N.C., 2021b. Metagenomic compendium of 189,680 DNA viruses from the human gut microbiome. Nat. Microbiol. 6, 960–970. https://doi.org/10.1038/s41564-021-00928-6

Nazir, A., Qi, C., Shi, N., Gao, X., Feng, Q., Qing, H., Li, F., Tong, Y., 2022. Characterization and Genomic Analysis of a Novel Drexlervirial Bacteriophage IME268 with Lytic Activity Against Klebsiella pneumoniae. Infect. Drug Resist. 15, 1533–1546. https://doi.org/10.2147/IDR.S347110

713 Newberry, F., Shibu, P., Smith-Zaitlik, T., Eladawy, M., McCartney, A.L., Hoyles, L., Negus, D.,
714 2023. Phage vB_KmiS-Kmi2C infects members of the Klebsiella oxytoca complex and
715 represents a novel genus of lytic bacteriophage. bioRxiv.
716 https://doi.org/10.1101/2023.01.12.523727
717 Nguyen, D.T., Lessor, L.E., Cahill, J.L., Rasche, E.S., Kuty Everett, G.F., 2015. Complete
718 Genome Sequence of Klebsiella pneumoniae Carbapenemase-Producing K.
719 pneumoniae Siphophage Sushi. Genome Announc. 3.
720 https://doi.org/10.1128/genomeA.00994-15
721 Nir-Paz, R., Gelman, D., Khouri, A., Sisson, B.M., Fackler, J., Alkalay-Oren, S., Khalifa, L.,
722 Rimon, A., Yerushalmy, O., Bader, R., Amit, S., Coppenhagen-Glazer, S., Henry, M.,
723 Quinones, J., Malagon, F., Biswas, B., Moses, A.E., Merril, G., Schooley, R.T.,
724 Brownstein, M.J., Weil, Y.A., Hazan, R., 2019. Successful Treatment of Antibiotic-
725 resistant, Poly-microbial Bone Infection With Bacteriophages and Antibiotics
726 Combination. Clin. Infect. Dis. Off. Publ. Infect. Dis. Soc. Am. 69, 2015–2018.
727 https://doi.org/10.1093/cid/ciz222
728 Nishimura, Y., Yoshida, T., Kuronishi, M., Uehara, H., Ogata, H., Goto, S., 2017. ViPTree: the
729 viral proteomic tree server. Bioinforma. Oxf. Engl. 33, 2379–2380.
730 https://doi.org/10.1093/bioinformatics/btx157
731 Pagès, H., Aboyoun, P., Gentleman, R., DebRoy, S., 2021. Biostrings: Efficient manipulation
732 of biological strings. R package version 2.62.0.
733 Paradis, E., Schliep, K., 2019. ape 5.0: an environment for modern phylogenetics and
734 evolutionary analyses in R. Bioinforma. Oxf. Engl. 35, 526–528.
735 https://doi.org/10.1093/bioinformatics/bty633
736 Park, E.-A., Kim, Y.-T., Cho, J.-H., Ryu, S., Lee, J.-H., 2017. Characterization and genome
737 analysis of novel bacteriophages infecting the opportunistic human pathogens
738 Klebsiella oxytoca and K. pneumoniae. Arch. Virol. 162, 1129–1139.
739 https://doi.org/10.1007/s00705-016-3202-3
740 Pertics, B.Z., Cox, A., Nyúl, A., Szamek, N., Kovács, T., Schneider, G., 2021. Isolation and
741 Characterization of a Novel Lytic Bacteriophage against the K2 Capsule-Expressing
742 Hypervirulent Klebsiella pneumoniae Strain 52145, and Identification of Its
743 Functional Depolymerase. Microorganisms 9, 650.
744 https://doi.org/10.3390/microorganisms9030650
745 Rangel-Pineros, G., Millard, A., Michniewski, S., Scanlan, D., Sirén, K., Reyes, A., Petersen, B.,
746 Clokie, M.R.J., Sicheritz-Pontén, T., 2021. From Trees to Clouds: PhageClouds for Fast
747 Comparison of ∼640,000 Phage Genomic Sequences and Host-Centric Visualization
748 Using Genomic Network Graphs. PHAGE 2, 194–203.
749 https://doi.org/10.1089/phage.2021.0008
750 Richardson, R.W., Lessor, L., O'Leary, C., Gill, J., Liu, M., 2019. Complete Genome Sequence
751 of Klebsiella pneumoniae Siphophage Sanco. Microbiol. Resour. Announc. 8.
752 https://doi.org/10.1128/MRA.01252-19
753 Rodríguez-Medina, N., Barrios-Camacho, H., Duran-Bedolla, J., Garza-Ramos, U., 2019.
754 Klebsiella variicola: an emerging pathogen in humans. Emerg. Microbes Infect. 8,
755 973–988. https://doi.org/10.1080/22221751.2019.1634981
756 Saldana, R., Newkirk, H., Liu, M., Gill, J.J., Ramsey, J., 2019. Complete Genome Sequence of
757 Shelby, a Siphophage Infecting Carbapenemase-Producing Klebsiella pneumoniae.
758 Microbiol. Resour. Announc. 8. https://doi.org/10.1128/MRA.01037-19

759 Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S.,
760 Rueden, C., Saalfeld, S., Schmid, B., Tinevez, J.-Y., White, D.J., Hartenstein, V., Eliceiri,
761 K., Tomancak, P., Cardona, A., 2012. Fiji: an open-source platform for biological-
762 image analysis. Nat. Methods 9, 676–682. https://doi.org/10.1038/nmeth.2019
763 Schliep, K., Potts, A.J., Morrison, D.A., Grimm, G.W., 2017. Intertwining phylogenetic trees
764 and networks. Methods Ecol. Evol. 8, 1212–1220. https://doi.org/10.1111/2041-
765 210X.12760
766 Scholl, D., Adhya, S., Merril, C., 2005. Escherichia coli K1's capsule is a barrier to
767 bacteriophage T7. Appl. Environ. Microbiol. 71, 4872–4874.
768 https://doi.org/10.1128/AEM.71.8.4872-4874.2005
769 Seemann, T., 2014. Prokka: rapid prokaryotic genome annotation. Bioinforma. Oxf. Engl. 30,
770 2068–2069. https://doi.org/10.1093/bioinformatics/btu153
771 Shibu, P., 2019. Investigations of carbapenem-resistant Klebsiella species and associated
772 clinical considerations. Westminster, London.
773 Shibu, P., McCuaig, F., McCartney, A.L., Kujawska, M., Hall, L.J., Hoyles, L., 2021. Improved
774 molecular characterization of the Klebsiella oxytoca complex reveals the prevalence
775 of the kleboxymycin biosynthetic gene cluster. Microb. Genomics 7.
776 https://doi.org/10.1099/mgen.0.000592
777 Smith-Zaitlik, T., Shibu, P., McCartney, A.L., Foster, G., Hoyles, L., Negus, D., 2022. Extended
778 genomic analyses of the broad-host-range phages vB_KmiM-2Di and vB_KmiM-4Dii
779 reveal slopekviruses have highly conserved genomes.
780 https://doi.org/10.1101/2022.04.06.486684
781 South, A., 2011. rworldmap: a New R package for mapping global data. R J. 3, 35–43.
782 Tabassum, R., Shafique, M., Khawaja, K.A., Alvi, I.A., Rehman, Y., Sheik, C.S., Abbas, Z.,
783 Rehman, S.U., 2018. Complete genome analysis of a Siphoviridae phage TSK1
784 showing biofilm removal potential against Klebsiella pneumoniae. Sci. Rep. 8, 17904.
785 https://doi.org/10.1038/s41598-018-36229-y
786 Taslem Mourosi, J., Awe, A., Guo, W., Batra, H., Ganesh, H., Wu, X., Zhu, J., 2022.
787 Understanding Bacteriophage Tail Fiber Interaction with Host Surface Receptor: The
788 Key "Blueprint" for Reprogramming Phage Host Range. Int. J. Mol. Sci. 23, 12146.
789 https://doi.org/10.3390/ijms232012146
790 Ter Horst, A.M., Santos-Medellín, C., Sorensen, J.W., Zinke, L.A., Wilson, R.M., Johnston,
791 E.R., Trubl, G., Pett-Ridge, J., Blazewicz, S.J., Hanson, P.J., Chanton, J.P., Schadt, C.W.,
792 Kostka, J.E., Emerson, J.B., 2021. Minnesota peat viromes reveal terrestrial and
793 aquatic niche partitioning for local and global viral populations. Microbiome 9, 233.
794 https://doi.org/10.1186/s40168-021-01156-0
795 Terzian, P., Olo Ndela, E., Galiez, C., Lossouarn, J., Pérez Bucio, R.E., Mom, R., Toussaint, A.,
796 Petit, M.-A., Enault, F., 2021. PHROG: families of prokaryotic virus proteins clustered
797 using remote homology. NAR Genomics Bioinforma. 3, lqab067.
798 https://doi.org/10.1093/nargab/lqab067
799 Thung, T.Y., White, M.E., Dai, W., Wilksch, J.J., Bamert, R.S., Rocker, A., Stubenrauch, C.J.,
800 Williams, D., Huang, C., Schittelhelm, R., Barr, J.J., Jameson, E., McGowan, S., Zhang,
801 Y., Wang, J., Dunstan, R.A., Lithgow, T., 2021. Component Parts of Bacteriophage
802 Virions Accurately Defined by a Machine-Learning Approach Built on Evolutionary
803 Features. mSystems 6, e0024221. https://doi.org/10.1128/mSystems.00242-21
804 Thurgood, T.L., Sharma, R., Call, J.J., Chronis, J.D., Dawson, D.D., Finnegan, Z.K., Foster, K.W.,
805 Meek, T., Potts, E., Sirrine, M.R., Atkinson, A.D., Fairholm, J.D., Handoko, Y.A., Ong,

K.L., Tateoka, O.B., Arens, D.K., Johnson, L., Kruger, J.L., Loertscher, E., Thompson, D.W., Walker, J.K., Robison, R.A., Casjens, S.R., Grose, J.H., 2020. Genome Sequences of 12 Phages That Infect Klebsiella pneumoniae. Microbiol. Resour. Announc. 9, e00024-20. https://doi.org/10.1128/MRA.00024-20

Tisza, M.J., Buck, C.B., 2021. A catalog of tens of thousands of viruses from human metagenomes reveals hidden associations with chronic diseases. Proc. Natl. Acad. Sci. U. S. A. 118, e2023202118. https://doi.org/10.1073/pnas.2023202118

Turner, D., Kropinski, A.M., Adriaenssens, E.M., 2021. A Roadmap for Genome-Based Phage Taxonomy. Viruses 13, 506. https://doi.org/10.3390/v13030506

Villarroel, J., Kleinheinz, K.A., Jurtz, V.I., Zschach, H., Lund, O., Nielsen, M., Larsen, M.V., 2016. HostPhinder: A Phage Host Prediction Tool. Viruses 8, 116. https://doi.org/10.3390/v8050116

Walker, P.J., Siddell, S.G., Lefkowitz, E.J., Mushegian, A.R., Adriaenssens, E.M., Alfenas-Zerbini, P., Davison, A.J., Dempsey, D.M., Dutilh, B.E., García, M.L., Harrach, B., Harrison, R.L., Hendrickson, R.C., Junglen, S., Knowles, N.J., Krupovic, M., Kuhn, J.H., Lambert, A.J., Łobocka, M., Nibert, M.L., Oksanen, H.M., Orton, R.J., Robertson, D.L., Rubino, L., Sabanadzovic, S., Simmonds, P., Smith, D.B., Suzuki, N., Van Dooerslaer, K., Vandamme, A.-M., Varsani, A., Zerbini, F.M., 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. https://doi.org/10.1007/s00705-021-05156-1

Walker, P.J., Siddell, S.G., Lefkowitz, E.J., Mushegian, A.R., Dempsey, D.M., Dutilh, B.E., Harrach, B., Harrison, R.L., Hendrickson, R.C., Junglen, S., Knowles, N.J., Kropinski, A.M., Krupovic, M., Kuhn, J.H., Nibert, M., Rubino, L., Sabanadzovic, S., Simmonds, P., Varsani, A., Zerbini, F.M., Davison, A.J., 2019. Changes to virus taxonomy and the International Code of Virus Classification and Nomenclature ratified by the International Committee on Taxonomy of Viruses (2019). Arch. Virol. 164, 2417–2429. https://doi.org/10.1007/s00705-019-04306-w

Wick, R.R., Heinz, E., Holt, K.E., Wyres, K.L., 2018. Kaptive Web: User-Friendly Capsule and Lipopolysaccharide Serotype Prediction for Klebsiella Genomes. J. Clin. Microbiol. 56. https://doi.org/10.1128/JCM.00197-18

Wick, R.R., Judd, L.M., Gorrie, C.L., Holt, K.E., 2017. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. PLoS Comput. Biol. 13, e1005595. https://doi.org/10.1371/journal.pcbi.1005595

Wick, R.R., Schultz, M.B., Zobel, J., Holt, K.E., 2015. Bandage: interactive visualization of de novo genome assemblies. Bioinforma. Oxf. Engl. 31, 3350–3352. https://doi.org/10.1093/bioinformatics/btv383

Wintachai, P., Naknaen, A., Thammaphet, J., Pomwised, R., Phaonakrop, N., Roytrakul, S., Smith, D.R., 2020. Characterization of extended-spectrum-β-lactamase producing Klebsiella pneumoniae phage KP1801 and evaluation of therapeutic efficacy in vitro and in vivo. Sci. Rep. 10, 11803. https://doi.org/10.1038/s41598-020-68702-y

Zurabov, F., Zhilenkov, E., 2019. Complete Genome Sequences of Lytic Polysaccharide-Degrading Klebsiella pneumoniae Bacteriophages vB_KpnS_FZ10, vB_KpnP_FZ12, vB_KpnM_FZ14, and vB_KpnS_FZ41. Microbiol. Resour. Announc. 8. https://doi.org/10.1128/MRA.00914-19

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

Strain *	Source	K type †	O antigen†	MLST ‡	Infection type §							
					KLPN1	KLPN2	KLPN3	KLPN4	KLPN5	KLPN6	KLPN7	KLPN8
L4-FAA5	Human caecum (Hoyles et al., 2015)	K2	O1v1	ST380	++++ d	++++ d	++++ d	++++ d	0	0	0	0
PS_misc2	Groin (Shibu, 2019)	K64	O1v1	ST14	0	0	0	0	0	0	0	0
PS_misc3	Rectum (Shibu, 2019)	K52	OL101	ST45	0	0	0	0	0	0	0	0
PS_misc5	Rectum (Shibu, 2019)	K81	OL101	16-24-21-31-47-22-281	0	0	0	0	++++ d	++++ d	++++ d	0
PS_misc6	Rectum (Shibu, 2019)	K110	O2v1	ST716	0	0	0	0	0	0	0	+++ d
PS_misc7	Rectum (Shibu, 2019)	K30	O1v1	ST294	0	0	0	0	0	0	0	0
PS_misc8	Perineum (Shibu, 2019)	K52	OL101	ST45	0	0	0	0	0	0	0	0
PS_Kpn1	Perineum (Shibu, 2019)	K2	O4	1-X-1-1-1-1-1	+++	++	++	++	0	0	0	0
PS_Kpn2	Rectum (Shibu, 2019)	K64	O1v1	ST14	+++	+++	+++	++	0	0	0	0
PS_Kpn3	Rectum (Shibu, 2019)	K64	O1v1	1-1-1-1-1-1-X	0	0	0	0	0	0	0	0
PS_Kpn4	Cross-infection (Shibu, 2019)	K64	O1v2	ST15	0	0	0	0	0	0	0	0
PS_Kpn7	Rectum (Shibu, 2019)	K2	O1v1	ST14	++	++	++	++	0	0	0	0
PS_Kpn9	Rectum (Shibu, 2019)	K18	O2v1	2-1-1-1-X-1-4	0	0	0	0	0	0	0	0
PS_Kpn10	Rectum (Shibu, 2019)	K2	O1v1	ST14	++++	+++	+++	++	0	0	0	0
PS_Kpn11	Urine (Shibu, 2019)	K107	O2v2	ST258	0	0	0	0	0	0	0	0
PS_Kpn12	Rectum (Shibu, 2019)	K15	O4	ST11	0	0	0	0	0	0	0	0
PS_Kpn13	Rectum (Shibu, 2019)	K2	O1v1	ST14	+++	+++	+++	0	0	0	0	0
PS_Kpn14	Rectum (Shibu, 2019)	K2	O1v1	ST14	++++	+++	+++	++	0	0	0	0
PS_Kpn15	Rectum (Shibu, 2019)	K52	OL101	ST45	0	0	0	0	0	0	0	0
PS_Kpn16	Rectum (Shibu, 2019)	K17	O1v1	ST101	0	0	0	0	0	0	0	0
PS_Kpn24	Rectum (Shibu, 2019)	K2	O1v1	ST14	++	++	++	++	++	++	++	0
PS_Kpn25	Rectum (Shibu, 2019)	K2	O1v1	ST14	+++	++	++	++	0	0	0	0
PS_Kpn26	Rectum (Shibu, 2019)	K64	O1v1	ST147	0	0	0	0	0	0	0	0
PS_Kpn27	Urine (Shibu, 2019)	K107	O2v2	ST258	0	0	0	0	0	0	0	0
PS_Kpn28	Urine (Shibu, 2019)	K2	O1v1	ST14	++++	+++	++	++	0	0	0	0
PS_Kpn29	Urine (Shibu, 2019)	K2	O1v1	ST14	++	++	++	++	0	0	0	0
PS_Kpn30	Mouth (Shibu, 2019)	K112	O1v1	ST15	0	0	0	0	0	0	0	0
PS_Kpn31	Perineum (Shibu, 2019)	K2	O1v1	ST14	++++	++	++	++	0	0	0	0
PS_Kpn32	Drain fluid (Shibu, 2019)	K22	O1v1	ST35	0	0	0	0	0	0	0	0
PS_Kpn33	Urine (Shibu, 2019)	K2	O1v1	ST14	+++ (d)	+++ (d)	++	++	0	0	0	0
PS_Kpn35	Urine (Shibu, 2019)	K2	O1v1	ST14	+++ (d)	+++ (d)	++	++	0	0	0	0
PS_Kpn36	Urine (Shibu, 2019)	K24	O2v1	ST11	0	0	0	0	0	0	0	0
PS_Kpn37	Wound (Shibu, 2019)	K51	O3b	ST16	0	0	0	0	0	0	0	0
PS_Kpn38	High vaginal swab (Shibu, 2019)	K2	O1v1	ST14	+++	++	+++	++	0	0	0	0
PS_Kpn39	Wound (Shibu, 2019)	K64	O1v1	ST14	0	0	0	0	0	0	0	0
PS_Kpn40	Wound (Shibu, 2019)	K107	O2v2	ST512	0	0	0	0	0	0	0	0
PS_Kpn41	Leg (Shibu, 2019)	K112	O1v1	ST15	0	0	0	0	0	0	0	0

853 * All strains with PS prefix identified as *K. pneumoniae* subsp. *pneumoniae* by average nucleotide identity and phylogenetic analyses against
854 type strains of the genus *Klebsiella*, except for PS_misc5 (*K. variicola* subsp. *variicola*) (Shibu, 2019).
855 † Determined using Kaptive (Wick et al., 2018).
856 ‡ MLST, multi-locus sequence type determined via PubMLST (Jolley et al., 2018).

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

861 **Table 2.** Webervirus genomes associated with *Klebsiella* hosts included in this study

Phage/species	Accession	Isolated from	Location	Reference
Klebsiella phage vB_KpnS-KLPN2	OM065837	Sewage	UK	This study
Klebsiella phage vB_KpnS-KLPN3	OM065838	Sewage	UK	This study
Klebsiella phage vB_KpnS-KLPN4	OM065839	Sewage	UK	This study
Klebsiella phage vB_KvaS-KLPN5	OM065840	Sewage	UK	This study
Klebsiella phage vB_KvaS-KLPN6	OM065841	Sewage	UK	This study
Klebsiella phage vB_KvaS-KLPN7	OM065842	Sewage	UK	This study
Klebsiella phage vB_KpnS-KLPN8	OM065843	Sewage	UK	This study
Klebsiella phage 066039 (P39)	MW042802.1	Sewage	China	(Fang et al., 2022)
Klebsiella phage ABTNL-2	MZ221764.1	Sewage	China	–
Klebsiella phage B1	MW672037.1	Sewage	Hungary	(Pertics et al., 2021)
Klebsiella phage IME268	MZ398242.1	Unknown	China	(Nazir et al., 2022)
Klebsiella phage LF20	MW417503.1	Hospital wastewater	China	–
Klebsiella phage MMBB	MT894005.1	Merri Creek watercourse	Australia	(Thung et al., 2021)
Klebsiella phage NJR15	NC_048044.1	Sewage	China	–
Klebsiella phage NJS2	NC_048043.1	Sewage	China	–
Klebsiella phage NJS3	MH633486.1	Hospital sewage	China	–
Klebsiella phage P1	MZ598515.1	Water	China	–
Klebsiella phage P528	MW021764.1	Sewage	USA	(Cranston et al., 2022)
Klebsiella phage PhiKpNIH-10	MN395285.1	Wastewater treatment plant	USA	(Hesse et al., 2020)
Klebsiella phage RAD2	NC_055956.1	Hospital wastewater	UK	(Dunstan et al., 2021)
Klebsiella phage Sanco	MK618657.1	Wastewater treatment plant	USA	(Richardson et al., 2019)
Klebsiella phage SH-Kp 160016	KY575286.1	Sewage	China	–
Klebsiella phage Solomon	MT701592.1	Wastewater treatment plant	USA	(Hudson et al., 2021)
Klebsiella phage vB_KpnS_2811	LR757892.1	Unknown	UK	(Haines et al., 2021)
Klebsiella phage vB_KpnS_KingDDD	MN013078.1	Sewage	USA	(Thurgood et al., 2020)
Klebsiella phage vB_KpnS_Penguinator	MN013087.1	Sewage	USA	(Thurgood et al., 2020)
Klebsiella phage vB_KpnS_ZX2	MW722081.1	Faeces	China	–
Klebsiella phage vB_kpnS-VAC10	MZ428227.1	Sewage	Spain	(Bleriot et al., 2021)
Klebsiella phage vB_KpnS-VAC11	MZ428228.1	Sewage	Spain	(Bleriot et al., 2021)
Klebsiella phage vB_kpnS-VAC2	MZ428221.1	Sewage	Spain	(Bleriot et al., 2021)
Klebsiella phage vB_KpnS-VAC4	MZ428222.1	Sewage	Spain	(Bleriot et al., 2021)
Klebsiella phage vB_KpnS-VAC5	MZ428223.1	Sewage	Spain	(Bleriot et al., 2021)
Klebsiella phage vB_KpnS-VAC6	MZ428224.1	Sewage	Spain	(Bleriot et al., 2021)
Klebsiella phage vB_KpnS-VAC7	MZ428225.1	Sewage	Spain	(Bleriot et al., 2021)
Klebsiella phage vB_KpnS-VAC8	MZ428226.1	Sewage	Spain	(Bleriot et al., 2021)
Webervirus alina	NC_049837.1	Sewage	USA	(Thurgood et al., 2020)
Webervirus call	NC_049836.1	Sewage	USA	(Thurgood et al., 2020)
Webervirus domnhall	NC_049835.1	Sewage	USA	(Thurgood et al., 2020)
Webervirus F20	NC_043469.1	Lake water	South Korea	(Mishra et al., 2012)
Webervirus FZ10	NC_049840.1	Sewage	Russia	(Zurabov and Zhilenkov, 2019)
Webervirus GHK3	NC_048162.1	Sewage	China	(Cai et al., 2019)
Webervirus IMGroot	NC_049834.1	Sewage	USA	(Thurgood et al., 2020)
Webervirus JY917	NC_049843.1	Unknown	China	–
Webervirus KL	NC_049838.1	Unknown	Unknown	–
Webervirus KLPN1	NC_028760.1	Human caecal effluent	UK	(Hoyles et al., 2015)
Webervirus KLPPOU149	NC_049842.1	Unknown	Cote d'Ivoire	–
Webervirus KOX1	NC_047825.1	Wastewater	Australia	(Brown et al., 2017)
Webervirus KP1801	NC_049848.1	Hospital wastewater	Thailand	(Wintachai et al., 2020)
Webervirus KP36	NC_029099.1	Wastewater	Poland	(Majkowska-Skrobek et al., 2016)
Webervirus KpCol1	NC_047907.1	Wastewater	Turkey	–
Webervirus KpKT21phi1	NC_048143.1	Unknown	USA	(Nir-Paz et al., 2019)
Webervirus KpNIH2	NC_049845.1	Wastewater	USA	(Hesse et al., 2020)
Webervirus KpV522	NC_047784.1	Sewage	Russia	–
Webervirus mezzogao	NC_047850.1	Wastewater	USA	(Gao et al., 2017)
Webervirus N141	NC_047841.1	Unknown	South Korea	–

Phage/species	Accession	Isolated from	Location	Reference
Webervirus NJS1	NC_048024.1	Sewage	China	–
Webervirus PKP126	NC_031053.1	Sewage	South Korea	(Park et al., 2017)
Webervirus segescirculi	NC_049833.1	Sewage	USA	(Thurgood et al., 2020)
Webervirus shelby	NC_049846.1	Pond water	USA	(Saldana et al., 2019)
Webervirus sin4	NC_049847.1	Wastewater	USA	(Castillo et al., 2019)
Webervirus skenny	NC_049841.1	Activated sludge	USA	(Gramer et al., 2019)
Webervirus sushi	NC_028774.1	Sewage	USA	(Nguyen et al., 2015)
Webervirus sweeny	NC_049839.1	Wastewater	USA	(Martinez et al., 2019)
Webervirus TAH8	NC_048042.1	Hospital sewage	China	–
Webervirus TSK1	NC_048126.1	Sewage	Pakistan	(Tabassum et al., 2018)
Webervirus wv13	NC_049844.1	Wastewater	Hungary	(Horváth et al., 2020)
Webervirus wv1513	NC_028786.1	Sewage	China	(Cao et al., 2015)

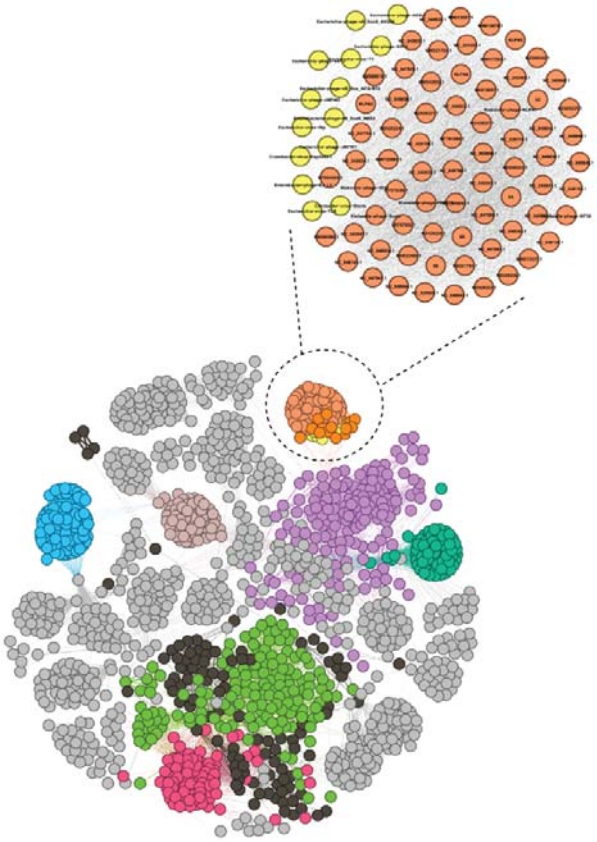
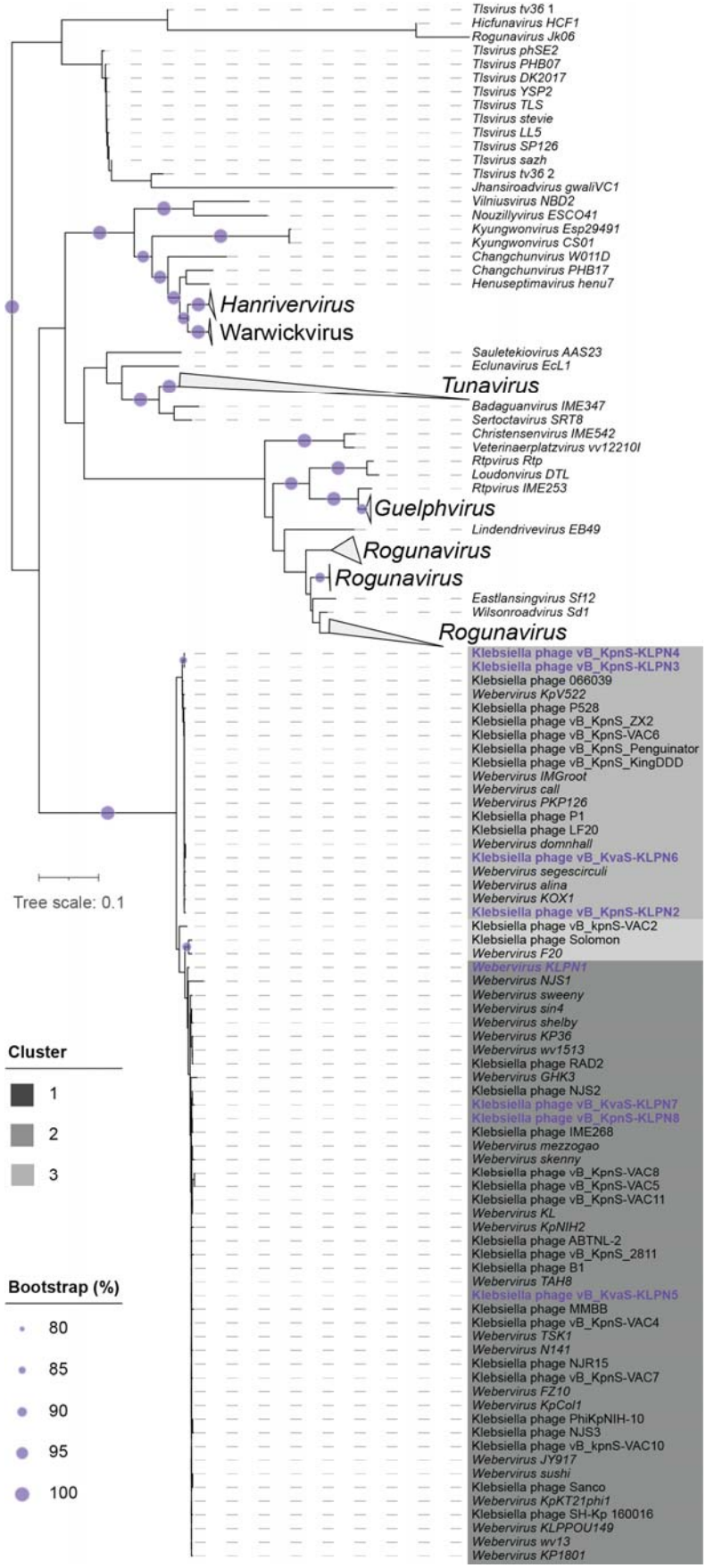


Figure 1. vConTACT filtered gene-sharing network in which only nodes connected to the main cluster are shown (filtered based on giant component; Fruchterman Reingold layout). The modularity of the network was determined (Gephi options selected: randomness, use weights, resolution 1.0). The network comprised 33 modules. Only modules representing >4 % of all nodes in the network are coloured. The circled module contained all 67 webervirus genomes (orange nodes) plus some other members of the family *Drexlerviridae* (yellow nodes). All the *Klebsiella* phage genomes (i.e. weberviruses) clustered together.


Tlsvirus tv36 1
Hicfunavirus HCF1
Rogunavirus Jk06
Tlsvirus phSE2
Tlsvirus PHB07
Tlsvirus DK2017
Tlsvirus YSP2
Tlsvirus TLS
Tlsvirus stevie
Tlsvirus LL5
Tlsvirus SP126
Tlsvirus sazh
Tlsvirus tv36 2
Jhansiroadvirus gwaliVC1
Vilniusvirus NBD2
Nouzillyvirus ESCO41
Kyungwonvirus Esp29491
Kyungwonvirus CS01
Changchunvirus W011D
Changchunvirus PHB17
Henuseptimavirus henu7
Hanrivervirus
Warwickvirus
Sauletekiovirus AAS23
Eclunavirus EcL1
Tunavirus
Badaguanvirus IME347
Sertoctavirus SRT8
Christensenvirus IME542
Veterinaerplatzvirus vv12210I
Rtpvirus Rtp
Loudonvirus DTL
Rtpvirus IME253
Guelphvirus
Lindendrivevirus EB49
Rogunavirus
Rogunavirus
Eastlansingvirus Sf12
Wilsonroadvirus Sd1
Rogunavirus
Klebsiella phage vB_KpnS-KLPN4
Klebsiella phage vB_KpnS-KLPN3
Klebsiella phage 066039
Webervirus KpV522
Klebsiella phage P528
Klebsiella phage vB_KpnS_ZX2
Klebsiella phage vB_KpnS-VAC6
Klebsiella phage vB_KpnS_Penguinator
Klebsiella phage vB_KpnS_KingDDD
Webervirus IMGroot
Webervirus call
Webervirus PKP126
Klebsiella phage P1
Klebsiella phage LF20
Webervirus domnhall
Klebsiella phage vB_KvaS-KLPN6
Webervirus segescirculi
Webervirus alina
Webervirus KOX1
Klebsiella phage vB_KpnS-KLPN2
Klebsiella phage vB_kpnS-VAC2
Klebsiella phage Solomon
Webervirus F20
Webervirus KLPN1
Webervirus NJS1
Webervirus sweeny
Webervirus sin4
Webervirus shelby
Webervirus KP36
Webervirus wv1513
Klebsiella phage RAD2
Webervirus GHK3
Klebsiella phage NJS2
Klebsiella phage vB_KvaS-KLPN7
Klebsiella phage vB_KpnS-KLPN8
Klebsiella phage IME268
Webervirus mezzogao
Webervirus skenny
Klebsiella phage vB_KpnS-VAC8
Klebsiella phage vB_KpnS-VAC5
Klebsiella phage vB_KpnS-VAC11
Webervirus KL
Webervirus KpNIH2
Klebsiella phage ABTNL-2
Klebsiella phage vB_KpnS_2811
Klebsiella phage B1
Webervirus TAH8
Klebsiella phage vB_KvaS-KLPN5
Klebsiella phage MMBB
Klebsiella phage vB_KpnS-VAC4
Webervirus TSK1
Webervirus N141
Klebsiella phage NJR15
Klebsiella phage vB_KpnS-VAC7
Webervirus FZ10
Webervirus KpCol1
Klebsiella phage PhiKpNIH-10
Klebsiella phage NJS3
Klebsiella phage vB_kpnS-VAC10
Webervirus JY917
Webervirus sushi
Klebsiella phage Sanco
Webervirus KpKT21phi1
Klebsiella phage SH-Kp 160016
Webervirus KLPPOU149
Webervirus wv13
Webervirus KP1801
Tree scale: 0.1
Cluster
1
2
3
Bootstrap (%)
80
85
90
95
100

874

Figure 2. Phylogenetic relationships (maximum-likelihood tree) of members of the family *Drexlerviridae* based on analysis of large-subunit terminase amino acid sequences encoded in phage genomes. Bootstrap values are expressed as a percentage of 100 replications; scale bar, mean number of amino acid substitutions per position; the tree is rooted at the midpoint. Clusters match those presented in **Supplementary Figure B** for the VIRIDIC analysis.

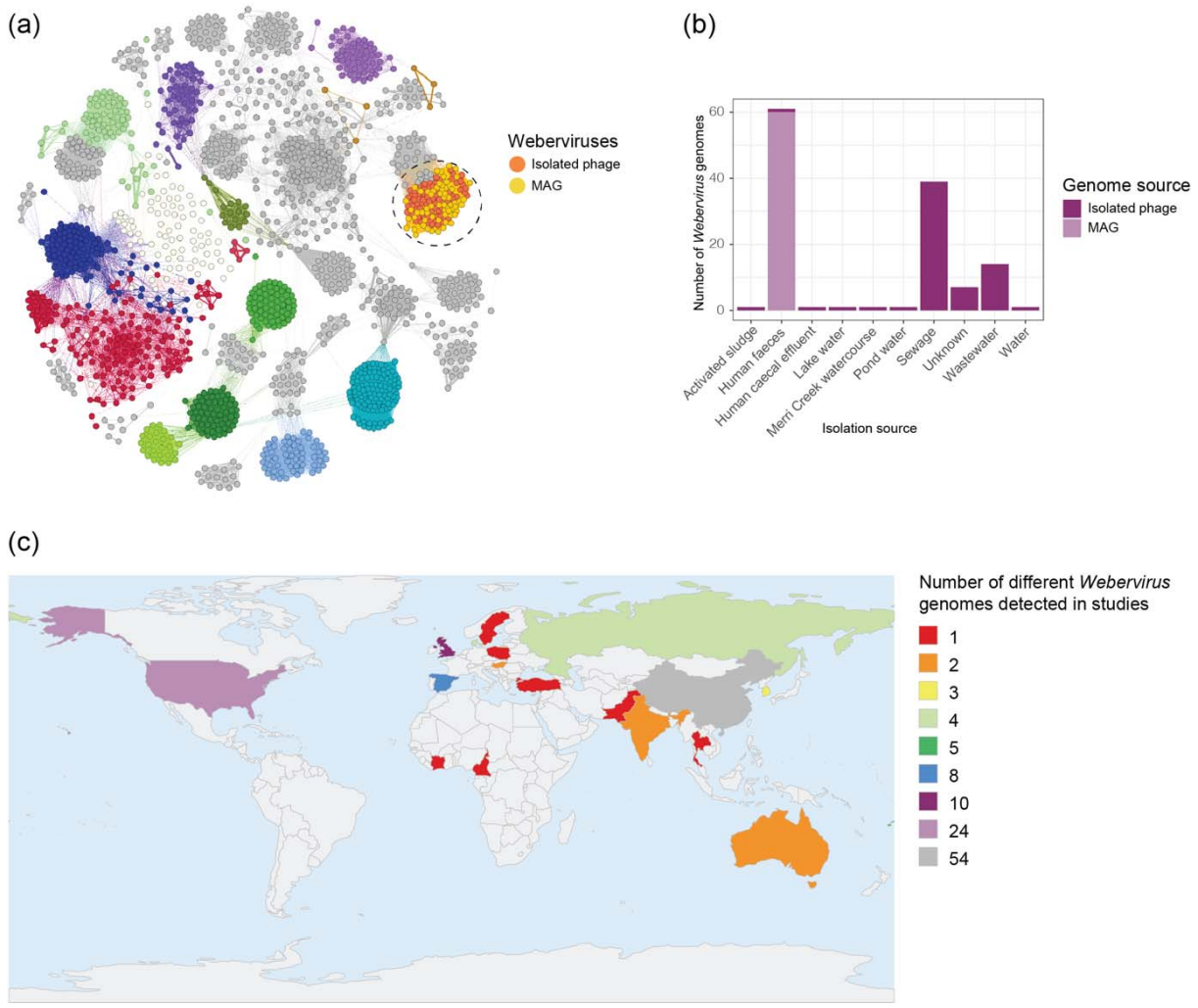
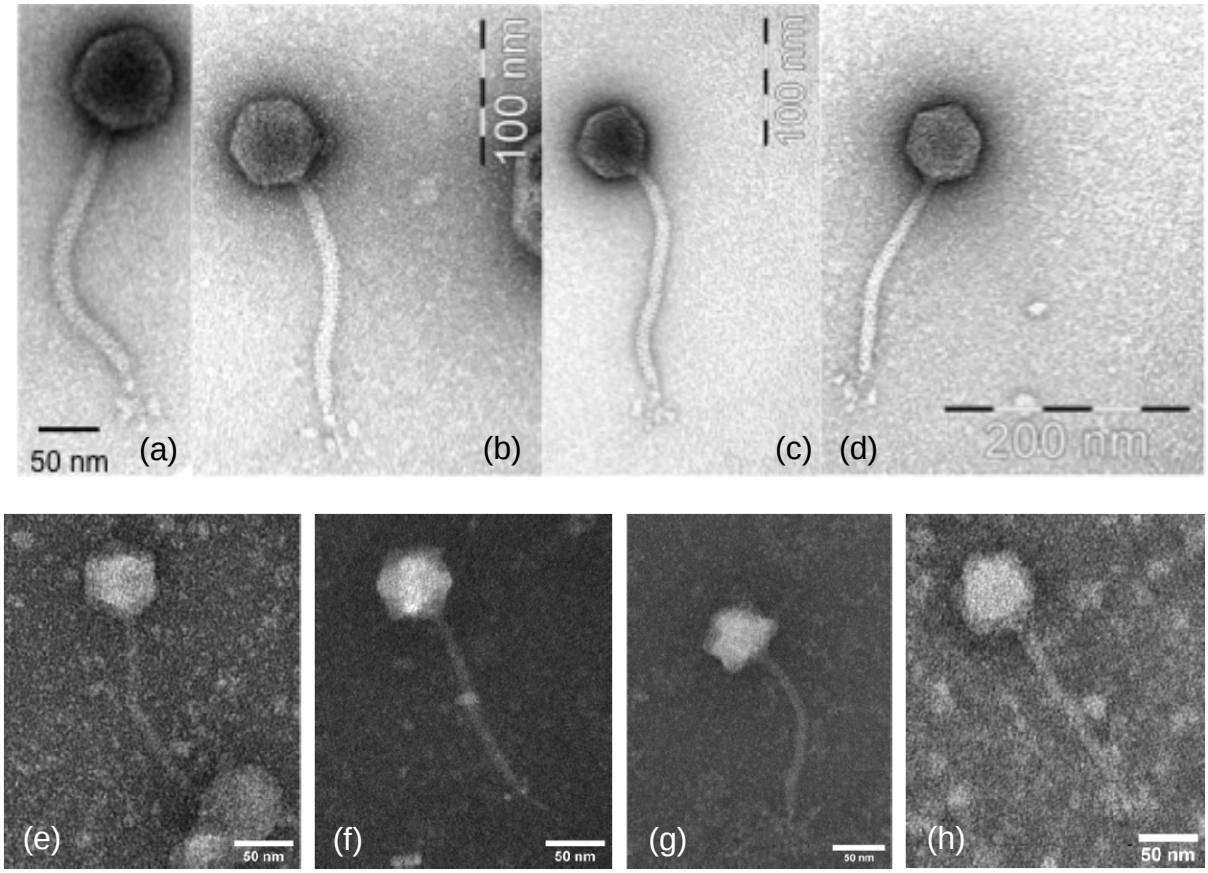
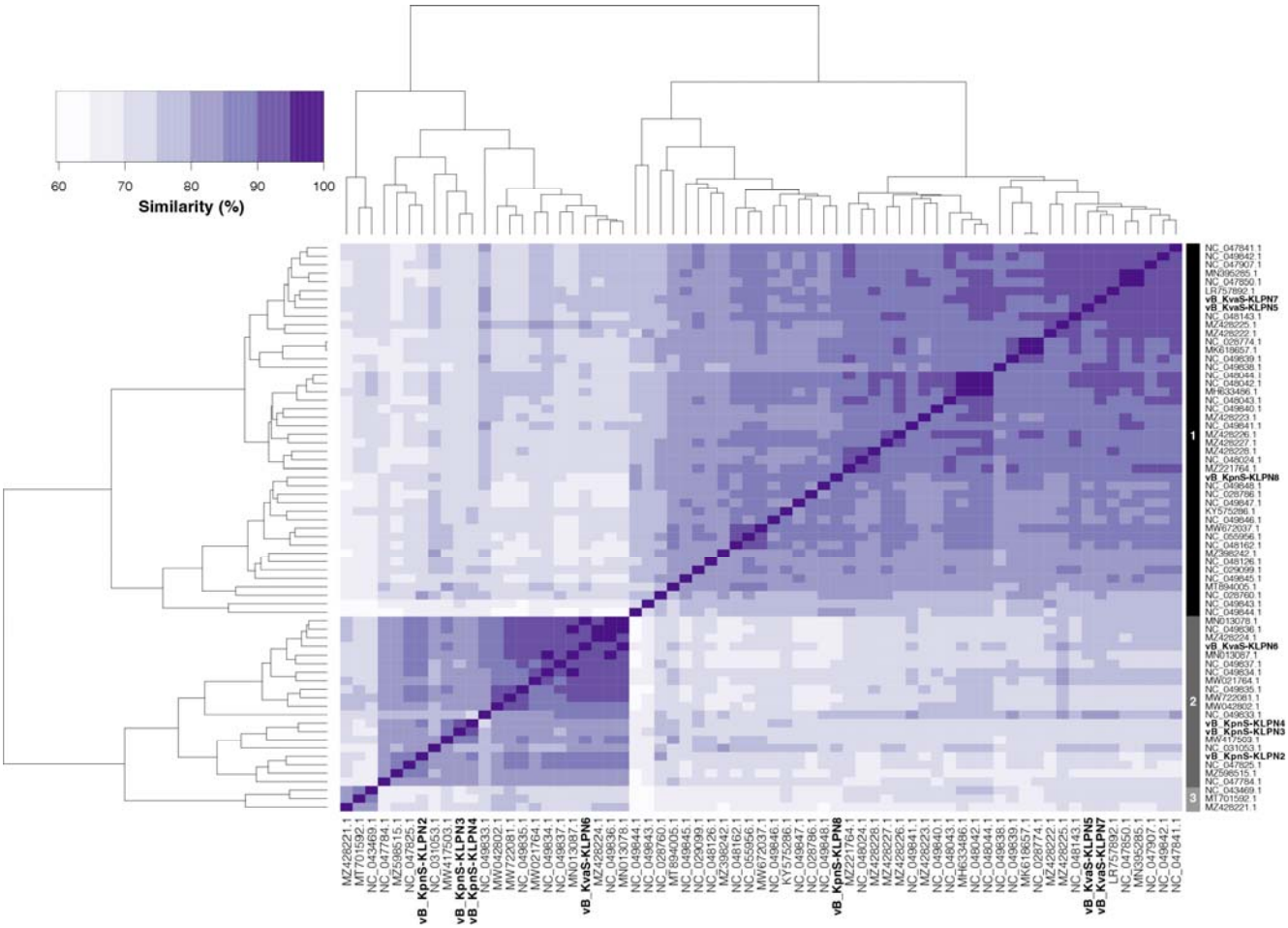


Figure 3. Distribution of weberviruses. (a) vConTACT filtered gene-sharing network in which only nodes connected to the main cluster are shown (Fruchterman Reingold layout; filtered based on giant component; Gephi modularity options selected – randomness, use weights, resolution 1.0). The network comprises 34 modules. Only modules representing >3 % of all nodes in the network are coloured. Module 0 (represented by the circled orange and yellow nodes in the image) contained only the 127 webervirus genomes (*n*=67 isolated phages; *n*=60 MAGs). (b) Stacked bar graph showing the sources of the 127 webervirus genomes (*n*=67 isolated phages; *n*=60 MAGs). (c) Geographical distribution of 126 of the webervirus genomes included in this study (the location information was not available for one isolated phage, namely *Webervirus KL*).

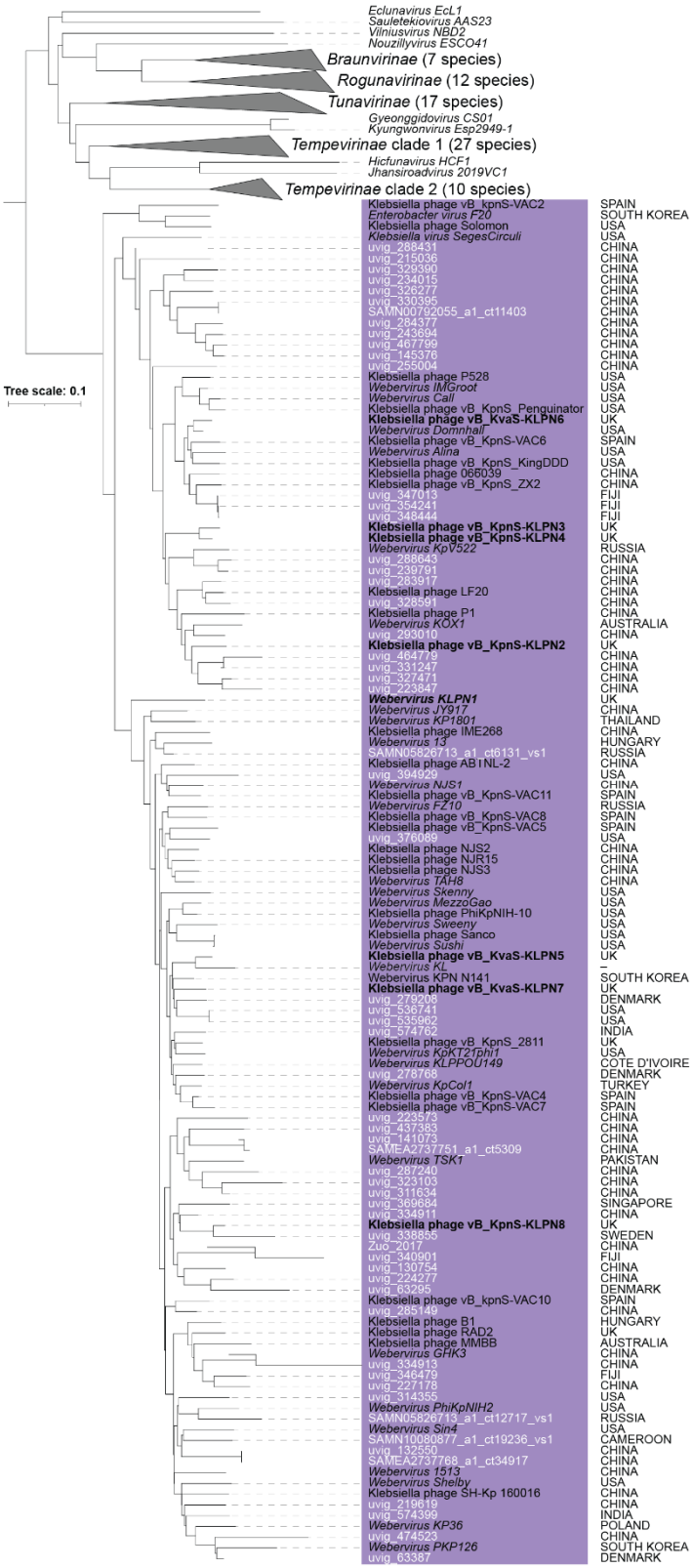
Supplementary figures for Dawson *et al.*, Weberviruses are gut-associated phages that infect *Klebsiella* spp.



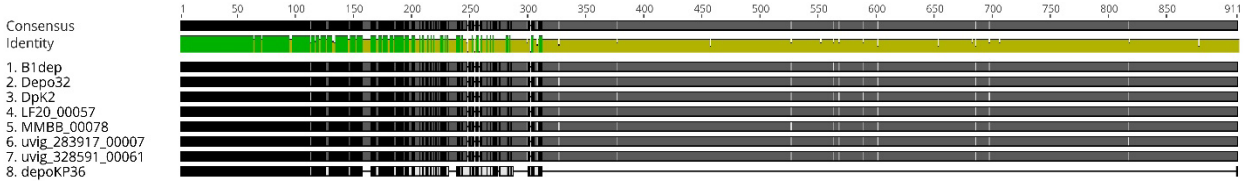
Supplementary Figure A. TEM images of (a) KLPN1, (b) vB_KpnS-KLPN2, (c) vB_KpnS-KLPN3, (d) vB_KpnS-KLPN4, (e) vB_KvaS-KLPN5, (f) vB_KvaS-KLPN6, (g) vB_KvaS-KLPN7 and (h) vB_KpnS-KLPN8.



Supplementary Figure B. Heatmap showing bidirectional clustering of VIRIDIC similarity data for the 67 webervirus genomes. The three numbered clusters at the side of the figure correspond to the three clusters detected by VIRIDIC. NC_028760.1 is the genome sequence for *Webervirus KLPN1*.



Supplementary Figure C. ViPTree analysis of 127 webervirus genomes with information (where known) provided for country of isolation of each phage.



Supplementary Figure D. MUSCLE amino acid alignment of depolymerases identified in this study. A high level of sequence conservation (indicated in green) was identified at the N-terminal region of all the depolymerase sequences (first 100 amino acids).