RESEARCH ARTICLE
Smith-Zaitlik *et al.*, *Microbiology* 2022;168:001247
DOI 10.1099/mic.0.001247

Extended genomic analyses of the broad-host-range phages vB_KmiM-2Di and vB_KmiM-4Dii reveal slopekviruses have highly conserved genomes

Thomas Smith-Zaitlik[1], Preetha Shibu[2]†, Anne L. McCartney[3], Geoffrey Foster[4], Lesley Hoyles[1,*] and David Negus[1,*]

Abstract

High levels of antimicrobial resistance among members of the *Klebsiella oxytoca* complex (KoC) have led to renewed interest in the use of bacteriophage (phage) therapy to tackle infections caused by these bacteria. In this study we characterized two lytic phages, vB_KmiM-2Di and vB_KmiM-4Dii, that were isolated from sewage water against two GES-5-positive *Klebsiella michiganensis* strains (PS_Koxy2 and PS_Koxy4, respectively). ViPTree analysis showed both phages belonged to the genus *Slopekvirus*. *rpoB* gene-based sequence analysis of 108 presumptive *K. oxytoca* isolates (*n*=59 clinical, *n*=49 veterinary) found *K. michiganensis* to be more prevalent (46% clinical and 43% veterinary, respectively) than *K. oxytoca* (40% clinical and 6% veterinary, respectively). Host range analysis against these 108 isolates found both vB_KmiM-2Di and vB_KmiM-4Dii showed broad lytic activity against KoC species. Several hypothetical homing endonuclease genes were encoded within the genomes of both phages, which may contribute to their broad host range. Differences in the tail fibre protein may explain the non-identical host range of the two phages. Pangenome analysis of 24 slopekviruses found that genomes within this genus are highly conserved, with more than 50% of all predicted coding sequences representing core genes at ≥95% identity and ≥70% coverage. Given their broad host ranges, our results suggest vB_KmiM-2Di and vB_KmiM-4Dii represent attractive potential therapeutics. In addition, current recommendations for phage-based pangenome analyses may require revision.

INTRODUCTION

Members of the *Klebsiella oxytoca* complex (KoC) are divided into phylogroups based on the sequence of their chromosomally encoded β-lactamase (bla_{OXY}) gene. The current phylogroups are *Klebsiella michiganensis* (KoI, with KoV sub-lineage), *K. oxytoca* (KoII), *K. spallanzanii* (KoIII), *K. pasteurii* (KoIV), *K. grimontii* (KoVI) and *K. huaxiensis* (KoVIII). KoVII has been described based on a single isolate [1–3].

Several members of the KoC can cause a variety of infections in humans including urinary tract infections, septicaemia and *Clostridioides*-negative antibiotic-associated haemorrhagic colitis [4–6]. The rapid development of antimicrobial resistance and the lack of novel antibiotics is a serious public health concern. Of 41 strains of the KoC isolated from bloodstream infections in the UK and Ireland, 100% were phenotypically resistant to amoxicillin and cefuroxime, 75.6% to piperacillin-tazobactam, 73.2% to amoxicillin-clavulanate and 48.8% to ciprofloxacin [7]. In a survey of 5724 clinical isolates of *K. oxytoca*, the SENTRY Antimicrobial Surveillance Programme identified rates of non-susceptibility of *K. oxytoca* to various antibiotics: 1.8% carbapenems,

Received 08 April 2022; Accepted 17 August 2022; Published 26 September 2022
Author affiliations: [1]Department of Biosciences, Nottingham Trent University, Nottingham NG1 4FQ, UK; [2]Life Sciences, University of Westminster, London, UK; [3]Department of Food and Nutritional Sciences, University of Reading, Reading, UK; [4]SRUC Veterinary Services, Inverness, UK.
***Correspondence:** Lesley Hoyles, lesley.hoyles@ntu.ac.uk; David Negus, david.negus@ntu.ac.uk
Keywords: homing endonuclease; *Klebsiella oxytoca complex*; lytic phage; myovirus; phage diversity; pangenome; *Slopekvirus*.
Abbreviations: ANI, average nucleotide identity; HEN, homing endonuclease; KoC, *Klebsiella oxytoca* complex; MAG, metagenome-assembled genome; MALDI-TOF MS, matrix-assisted laser desorption/ionisation-time of flight mass spectrometry; MLST, multi-locus sequence typing; MSA, multiple-sequence alignment; NA, nutrient agar; NB, nutrient broth; SD, standard deviation; TEM, transmission electron microscopy.
†Present address: Berkshire and Surrey Pathology Services, Frimley Health NHS Trust, Wexham Park Hospital, Slough, UK.
Repositories: The sequences for the two phage genomes described herein have been deposited in DDBJ/ENA/GenBank under accession numbers MZ707156 (vB_KmiM-2Di) and MZ707157 (vB_KmiM-4Dii) (BioProject PRJNA750911). Supplementary tables and data associated with this article are available from figshare at https://figshare.com/projects/Extended_genomic_analyses_of_the_broad-host range_phages_vB_KmiM-2Di_and_vB_KmiM 4Dii_reveal_slopekviruses_have_highly_conserved_genomes/127976.
Eight supplementary figures and five supplementary tables are available with the online version of this article.

12.5% ceftriaxone, 7.1% ciprofloxacin, 0.8% colistin and 0.1% tigecycline [8]. GES-positive clinical strains of the KoC have also been identified recently [9, 10].

KoC bacteria can also cause disease in animals. Of 336 samples collected from companion animals, 11 (3.3%) isolates were identified as *K. oxytoca*. These were typically recovered from the urogenital system and 81.8% were resistant to ampicillin [11]. *Klebsiella* spp. were detected in 51/1541 (3.3%) equine samples. Two *K. michiganensis* and one *K. oxytoca* were identified as non-repetitive cefotaxime-resistant isolates. These three isolates were phenotypically resistant to gentamicin, tobramycin, tetracycline, doxycycline, chloramphenicol and trimethoprim/sulfamethoxazole [12].

Accurate identification at the species level is important for recognizing the epidemiological and clinical significance of each member of the KoC in both humans and animals. Current diagnostics are unable to consistently differentiate between members of the KoC leading to historic misidentification of non-*K. oxytoca* strains as *K. oxytoca* [10]. The phenotypic similarity of the complex members prevents accurate identification using biochemical tests such as API 20E tests [1]. While MALDI-TOF MS is a rapid and cost-effective diagnostic tool, many databases in clinical and veterinary use have not been updated and lack biomarkers required to consistently differentiate members of the KoC [13, 14]. The use of 16S rRNA gene sequencing is considered unsuitable for identifying *Klebsiella* at the species level as the gene sequence is so highly conserved in this taxon [15]. *gyrA* and *rpoB* gene sequences have been used to distinguish *K. grimontii* and *K. huaxiensis* from other members of the KoC, supported by results from average nucleotide identity (ANI) analysis of genome sequence data [16, 17]. More recently, *rpoB* gene sequence analysis has been used to identify KoC, *Klebsiella pneumoniae* complex and *Raoultella* spp. isolates recovered from the faeces of healthy women and breast-fed infants [18]. MLST can also be used to assign strains to species of the KoC [10].

Bacteriophages (phages) are promising alternatives or adjuncts to current antibiotic therapies. We previously published an extensive review on *Klebsiella* phage and their potential as therapeutics [19]. The majority of *Klebsiella* phage publications focus on *K. pneumoniae*. Phage vB_Kox_ZX8 was isolated from human faeces; this phage was shown to clear bacteraemia caused by a clinical strain of *K. oxytoca* in BALB/c mice [20]. Another recent study described 30 novel phages that were active against *Klebsiella* species including KoC members *K. oxytoca* and *K. michiganensis* [21]. Of the phages isolated, 15 were active against the single *K. michiganensis* strain tested, whereas 16 showed activity against at least one of the five *K. oxytoca* strains tested. More recently, the lytic *Drexlerviridae* phage KMI8 was isolated against *K. michiganensis* [22]. KMI8 was lytic against 3/5 *K. michiganensis* strains but not *K. pneumoniae* (0/5) or *K. oxytoca* (0/1). To the best of our knowledge, ISF3 and ISF6 [23] and RP180 [24] are the only three *Raoultella* phages recorded in the literature. They were isolated against *Raoultella ornithinolytica* and *Raoultella* spp., respectively.

This study aimed to characterize the morphology, genomes and host ranges of two lytic phages, vB_KmiM-2Di and vB_KmiM-4Dii, isolated against two strains of GES-5-encoding *K. michiganensis* (PS_Koxy2 and PS_Koxy4, respectively) [25]. We used *rpoB* gene sequence analysis to accurately identify 108 clinical and veterinary isolates identified as *K. oxytoca* using MALDI-TOF MS and/or API 20E. These isolates were used in our host range analysis, to determine the therapeutic potential of vB_KmiM-2Di and vB_KmiM-4Dii against *Klebsiella* spp. A pangenome analysis was undertaken to compare our new phage genomes with those of their closest relatives.

METHODS

Strain information

Details of all strains included in this study can be found in Table 1.

Isolation of lytic phage

Filter-sterilized sewage samples (0.45 µm pore size cellulose acetate filter; Millipore) collected from mixed-liquor tanks at Mogden Sewage Treatment Works (March 2017) were screened against *K. michiganensis* strains PS_Koxy2 and PS_Koxy4 [10, 25]. Firstly, 9 ml of filter-sterilized sewage were added to 1 ml of 10×concentrated sterile nutrient broth (NB) (Oxoid) containing 50 mM $CaCl_2$ and 50 mM $MgCl_2$. This was then inoculated with 200 µl of overnight culture from each strain and incubated for 6 h at 37 °C. The samples were centrifuged at 10000 r.p.m. for 5 min; the supernatants were aliquoted (200 µl) and used in spot assays to identify lytic phage. Plaques were propagated to purity to create phage stocks.

Growth media and culture conditions

Bacterial cultures were initially streaked onto MacConkey agar (Sigma Aldrich) to ensure purity before being grown on nutrient agar (NA) (Sigma Aldrich). NB (Sigma Aldrich) was used for overnight cultures, incubated aerobically at 37 °C. All media used for phage assays were supplemented with $CaCl_2$ and $MgCl_2$ (final concentration 0.5 mM) unless otherwise specified.

Table 1. Clinical and veterinary isolates included in this study

With the exceptions of Ko8, Ko31 and GFKo17 (*K. oxytoca/R. ornithinolytica*), all strains were initially identified as *K. oxytoca*

Lab ID*	Source	Method of initial identification†	*rpoB* gene-based identity (sequence similarity, %)
Ko1	Blood culture (hand)	M	*K. oxytoca* ATCC 13182^T (99.80)
Ko2	Blood culture	M	*K. oxytoca* ATCC 13182^T (99.80)
Ko3	Sputum	M	*K. michiganensis* W14^T (99.60)
Ko4	Blood culture	M	*K. oxytoca* ATCC 13182^T (99.80)
Ko5	Blood culture (peripheral blood)	M	*K. michiganensis* W14^T (99.60)
Ko6	Urine	M	*K. oxytoca* ATCC 13182^T (99.60)
Ko7	Blood culture (arterial line)	M	*K. oxytoca* ATCC 13182^T (99.80)
Ko8	Blood culture	M	*K. grimontii* 06D021^T (99.40)
Ko9	Blood culture (peripheral blood)	M	*K. grimontii* 06D021^T (99.40)
Ko10	Blood culture	M	*K. michiganensis* W14^T (99.40)
Ko11	Blood culture (white waste)	M	*K. oxytoca* ATCC 13182^T (99.80)
Ko12	Blood culture	M	*K. oxytoca* ATCC 13182^T (99.80)
Ko13	Blood culture	M	*K. michiganensis* W14^T (99.60)
Ko14	Blood culture (peripheral blood)	M	*K. michiganensis* W14^T (99.40)
Ko15	Tissue (toe)	M	*K. oxytoca* ATCC 13182^T (99.60)
Ko16	Tissue (hip)	M	*K. grimontii* 06D021^T (99.60)
Ko17	Peritoneal fluid	M	*K. oxytoca* ATCC 13182^T (99.40)
Ko18	Sputum	M	*K. michiganensis* W14^T (99.40)
Ko19	Blood culture (peripheral blood)	M	*K. oxytoca* ATCC 13182^T (100.00)
Ko20	Blood culture	M	*K. oxytoca* ATCC 13182^T (99.60)
Ko21	Perfusion fluid	M	*K. michiganensis* W14^T (99.60)
Ko22	Perfusion fluid	M	*K. michiganensis* W14^T (99.60)
Ko23	Blood culture (venous blood)	M	*K. michiganensis* W14^T (99.60)
Ko24	Blood culture (systemic)	M	*K. michiganensis* W14^T (99.57)
Ko25	Clean catch urine	M	*K. oxytoca* ATCC 13182^T (99.80)
Ko26	Urine	M	*K. oxytoca* ATCC 13182^T (99.80)
Ko27	Brocho-alveolar lavage	M	*K. grimontii* 06D021^T (99.60)
Ko28	Blood culture (peripheral blood)	M	*K. michiganensis* W14^T (99.60)
Ko29	Sputum	M	*K. michiganensis* W14^T (99.40)
Ko30	Peritoneal fluid	M	*K. grimontii* 06D021^T (99.35)
Ko31	Sputum	M	*K. michiganensis* W14^T (99.40)
Ko32	Blood culture (venous blood)	M	*K. michiganensis* W14^T (99.60)
Ko33	Sputum	M	*K. michiganensis* W14^T (99.35)
Ko34	Blood culture (venous blood)	M	*K. pneumoniae* ATCC 13883^T (99.80)
Ko35	Sputum	M	*K. michiganensis* F107 CP024643.1 (97.67) ‡
Ko36	Blood culture (Hickman line)	M	*K. michiganensis* W14^T (99.60)
Ko37	Blood culture (venous blood)	M	*K. oxytoca* ATCC 13182^T (99.80)
Ko38	Blood culture (venous blood)	M	*K. oxytoca* ATCC 13182^T (99.80)

Continued

Table 1. Continued

Lab ID*	Source	Method of initial identification†	*rpoB* gene-based identity (sequence similarity, %)
Ko39	Blood culture (CVC line)	M	*K. michiganensis* W14^T (99.60)
Ko40	Blood culture (arm)	M	*K. oxytoca* ATCC 13182^T (99.60)
Ko41	Blood culture (peripheral blood)	M	*K. michiganensis* W14^T (99.20)
Ko42	Peritoneal fluid	M	*K. grimontii* 06D021^T (99.60)
Ko43	Bile fluid	M	*K. michiganensis* W14^T (99.80)
Ko44	CAPD fluid	M	*K. oxytoca* ATCC 13182^T (100.00)
Ko45	Aspirate (liver)	M	*K. oxytoca* ATCC 13182^T (99.35)
Ko46	Blood culture (peripheral blood)	M	*K. michiganensis* W14^T (100.00)
Ko47	Blood culture (central line)	M	*K. oxytoca* ATCC 13182^T (99.80)
Ko48	Blood culture (ACF blood)	M	*K. oxytoca* ATCC 13182^T (99.80)
Ko49	Tissue	M	*K. michiganensis* W14^T (100.00)
Ko50	Bone biopsy (heel)	M	*K. michiganensis* W14^T (99.20)
Ko51	Fluid (pancreas)	M	*K. grimontii* 06D021^T (99.40)
Ko52	Blood culture (peripheral blood)	M	*K. michiganensis* W14^T (99.60)
Ko53	Blood culture	M	*K. oxytoca* ATCC 13182^T (99.60)
Ko54	Blood culture (peripheral blood)	M	*K. oxytoca* ATCC 13182^T (99.78)
Ko55	Mid-stream urine	M	*K. oxytoca* ATCC 13182^T (99.80)
Ko56	Blood culture (peripheral blood)	M	*K. michiganensis* W14^T (99.60)
Ko57	Sputum	M	*K. oxytoca* ATCC 13182^T (99.80)
Ko58	Perfusion fluid	M	*K. michiganensis* W14^T (100.00)
Ko59	Blood culture (venous blood)	M	*K. michiganensis* W14^T (99.40)
GFKo2	Seal, foetal stomach; Rosshire	A	*K. grimontii* 06D021^T (99.40)
GFKo3	Common seal, peritoneal fluid; Angus	A	*K. grimontii* 06D021^T (99.40)
GFKo4	Common seal, lung; Inverness-shire	A	*R. ornithinolytica* NBRC 105727^T (99.60)
GFKo5	Sparrow, eye; England	A	*K. grimontii* 06D021^T (99.60)
GFKo6	African elephant, lung; Perthshire	A	*K. oxytoca* ATCC 13182^T (99.80)
GFKo7	Porpoise, SI; Berwickshire	A	*K. michiganensis* W14^T (99.60)
GFKo8	Porpoise, lung; Berwickshire	A	*K. michiganensis* W14^T (99.60)
GFKo9	Porpoise, lung; Perthshire	A	*R. ornithinolytica* NBRC 105727^T (99.80)
GFKo10	Bottlenose dolphin, intestine; Ross-shire	A	*K. michiganensis* W14^T (99.00)
GFKo11	Bovine, Milk, Perthshire	M	*K. huaxiensis* WCHKl090001^T (100.00)
GFKo12	Bovine, Milk, Not known	A	*K. grimontii* 06D021^T (99.40)
GFKo13	Bovine, Milk, Dumfries	M	*K. michiganensis* W14^T (99.60)
GFKo14	Bovine, Milk, Dumfries	M	*K. michiganensis* W14^T (99.40)
GFKo15	Bovine, Milk, England	M	*K. michiganensis* W14^T (99.40)
GFKo16	Bovine, Milk, England	M	*K. oxytoca* ATCC 13182^T (99.60)
GFKo17	Bovine, Milk sock, Fife	M	*R. ornithinolytica* NBRC 105727^T (99.40)
GFKo18	Bovine, Milk, England	A	*K. michiganensis* W14^T (99.40)
GFKo19	Bovine, Foetal stomach, England	M	*K. michiganensis* W14^T (99.40)
GFKo20	Bovine, Foetal stomach, England	A	*K. grimontii* 06D021^T (99.60)

Continued

Table 1. Continued

Lab ID*	Source	Method of initial identification†	*rpoB* gene-based identity (sequence similarity, %)
GFKo21	Bovine, Milk, Midlothian	A	*K. grimontii* 06D021^T (99.40)
GFKo22	Bovine, Milk, England	M	*K. grimontii* 06D021^T (99.60)
GFKo24	Bovine, Milk, England	M	*R. ornithinolytica* NBRC 105727^T (99.80)
GFKo25	Bovine, Milk, England	A	*R. ornithinolytica* NBRC 105727^T (99.80)
GFKo26	Bovine, Lung, Caithness	A	*K. grimontii* 06D021^T (99.60)
GFKo27	Bovine, Milk, unknown	M	*K. grimontii* 06D021^T (99.60)
GFKo28	Bovine, Milk, Wales?	M	*K. michiganensis* W14^T (99.40)
GFKo29	Bovine, Milk, Lanark	M	*K. michiganensis* W14^T (99.40)
GFKo30	Bovine, Milk, unknown	M	*K. michiganensis* W14^T (100.00)
GFKo31	Bovine, Milk sock, Moray	A	*K. michiganensis* W14^T (99.20)
GFKo32	Canine, Urine, Aberdeen	A	*K. michiganensis* W14^T (99.40)
GFKo33	Bovine, Lung, Orkney	A	*K. grimontii* 06D021^T (99.40)
GFKo34	Gecko, Skin, Caithness	A	*K. grimontii* 06D021^T (99.60)
GFKo35	Bovine, Spleen, Orkney	A	*K. michiganensis* W14^T (99.40)
GFKo36	Red deer, Lung, Orkney	A	*K. michiganensis* W14^T (99.40)
GFKo37	Bovine, Milk, Orkney	A	*K. michiganensis* W14^T (99.59)
GFKo38	Bovine, Milk, Dumfries	A	*K. michiganensis* W14^T (99.35)
GFKo39	Bovine, Milk, Dumfries	A	*K. grimontii* 06D021^T (99.40)
GFKo40	Bovine, Milk, Dumfries	A	*K. michiganensis* W14^T (99.60)
GFKo41	Bovine, Milk, Dumfries	A	*K. michiganensis* W14^T (99.60)
GFKo42	Bovine, Milk, Wigtown	A	*K. pneumoniae* ATCC 13883^T (99.60)
GFKo43	Bovine, Milk, Dumfries	A	*R. ornithinolytica* NBRC 105727^T (99.80)
GFKo44	Bovine, Milk, Dumfries	A	*K. michiganensis* W14^T (99.60)
GFKo45	Bovine, Milk, Dumfries	A	*K. michiganensis* W14^T (99.56)
GFKo46	Poultry layer, Liver, Lanark	Unknown	*K. pneumoniae* ATCC 13883^T (99.60)
GFKo47	Canine, Nose, Peebles	Unknown	*K. oxytoca* ATCC 13182^T (99.80)
GFKo48	Bovine, Foetal stomach, Ayr	Unknown	*K. pneumoniae* ATCC 13883^T (99.80)
GFKo49	Partridge, Liver, Wigtown	Unknown	*K. pneumoniae* ATCC 13883^T (99.60)
GFKo50	Bovine, Milk, Lanark	Unknown	*K. huaxiensis* WCHKl090001T (99.40)
GFKo52	Bovine, Milk, Ayr	Unknown	*R. terrigena* NBRC 14941^T (100.00)

*Ko prefix, clinical isolate (all obtained from the Pathogen Bank of Queen's Medical Centre Nottingham); GFKo prefix, veterinary isolate.
†A, API 20E; M, MALDI-TOF MS.
‡Identified through BLASTn analysis.

Colony PCR and sequencing of *rpoB* gene products

Forward (5′-GTTTTCCCAGTCACGACGTTGTAGGCGAAATGGCGGAAAACCA-3′) and reverse (5′-TTGTGAGC GGATAACAATTTCGAGTCTTCGAAGTTGTAACC-3′) *rpoB*-specific primers (Macrogen) [26] were diluted to 10 µM in DNase-free H_2O. A single colony for each isolate was touched using a sterile loop and dipped into the PCR master mix: 25 µl MangoMix (Meridian Bioscience); forward and reverse primers (0.5 µM final concentration); 20 µl DNase-free H_2O. DNA loading dye was included in the MangoMix. The positive control tube contained 2 µl of concentrated *K. michiganensis* PS_Koxy1 DNA [10] and the negative control contained master mix alone (no template DNA). The cycle conditions were as follows: initial denaturation, 95 °C for 10 min; 35 cycles 95 °C for 30 s, 54 °C for 30 s, 72 °C for 1 min; final extension, 72 °C for

5 min. PCR products were checked for single bands of expected size (1076 nt) using agarose gel electrophoresis (1% agarose gel in 1×TAE buffer; 100 V, 40 min) against a GeneRuler 1 kb ladder (ThermoFisher Scientific).

The Thermo Scientific Gene JET PCR Purification Kit was used to clean PCR products. Purified samples were checked for DNA concentration and purity using a NanoDrop 2000 spectrophotometer. Samples were then adjusted to 10 ng μl^{-1} and sent for sequencing (Source BioScience) using the MLST forward primer (5′-GTTTTCCCAGTCACGACGTTGTA-3′) [26].

Phylogenetic analysis of *rpoB* gene sequences

Returned *rpoB* gene sequences were trimmed to 501 nt using Geneious Prime (v2020.0.5) by extracting the sequence between 276 and 776 nt, and with reference to the 45 *rpoB* allele sequences (released 19 June 2020) available for download from the PubMLST *Klebsiella oxytoca*/*michiganensis*/*grimontii*/*pasteurii* typing database [27]. *rpoB* gene sequences were extracted from the following genomes and used in analyses: *K. oxytoca* (GCA_900977765), *K. spallanzanii* (GCA_901563875), *K. pasteurii* (GCA_901563825), *K. grimontii* (GCA_900200035), *K. michiganensis* (GCA_901556995), *K. pneumoniae* subsp. *ozaenae* (GCA_000826585), *K. pneumoniae* subsp. *rhinoscleromatis* (GCA_000163455), *K. pneumoniae* subsp. *pneumoniae* (GCA_000742135), *K. quasipneumoniae* subsp. *similipneumoniae* (GCA_900978135), *K. quasipneumoniae* subsp. *quasipneumoniae* (GCA_000751755), *K. africana* (GCA_900978845), '*K. quasivariicola*' (GCA_000523395), *K. variicola* subsp. *tropica* (GCA_900978675), *K. variicola* subsp. *variicola* (GCA_900977835), *K. aerogenes* (GCA_003417445), *K. indica* (GCA_005860775) and *K. huaxiensis* (GCA_003261575). *Raoultella electrica* (GCA_006711645), *R. terrigena* (GCA_006539725), *R. planticola* (GCA_000735435) and *R. ornithinolytica* (GCA_001598295) *rpoB* gene sequences were also included in analyses, as these taxa should be classified as *Klebsiella* spp. [28]. A multiple-sequence alignment (MSA; available as Supplementary Material, available with the online version of this article) was created using Clustal Omega (v1.2.2). The Jukes–Cantor genetic distance model was used to generate a neighbour-joining tree using the *rpoB* gene sequence of *K. aerogenes* ATCC 13048^T as an outgroup. The resulting newick file (available from figshare) was exported to iTOL (v6.1.1) [29] for visualization and annotation of the phylogenetic tree.

Host range analysis

Sterile molten top NA (3 ml; 0.2% SeaPlaque Agarose, Lonza) supplemented with $CaCl_2$ and $MgCl_2$ (both at 5 mM) was aliquoted into sterile test tubes held at 45 °C. Each tube was then inoculated with 250 µl of an overnight culture of the prospective host strain, and gently swirled to mix the contents before being poured onto an NA plate. The plate was gently swirled to ensure even distribution of top agar. Once set, 5 µl aliquots of both phages were spotted onto the plate (vB_KmiM-2Di, 1.67×10^9 pfu/ml; vB_KmiM-4Dii, 1.87×10^9 pfu/ml). Plates were incubated overnight at 37 °C. Next day, plates were inspected for lysis, with results recorded according to a modification of Haines *et al.* [30]: ++, complete lysis; +, hazy lysis; 0, no visible plaques. We also noted whether depolymerase activity (*d*) was evident (i.e. formation of haloes around plaques).

Phage concentration

The Vivaspin 20 50 kDa centrifugal concentrator (Cytiva) was used to concentrate 20 ml of filter-sterilized propagated phage. Samples were spun at 3000 ***g*** until only 200 µl of the sample remained. This concentrated phage stock was stored at 4 °C.

Transmission electron microscopy (TEM)

Formvar/carbon-coated 200 mesh copper grids (Agar Scientific) were prepared via glow discharge (10 mA, 10 s) using a Q150R ES sputter coater (Quorum Technologies). Phage suspensions (15 µl) were pipetted onto the grid surface for 30 s before removal using filter paper. Samples were stained using 15 µl of 2% phosphotungstic acid. Excess stain was removed using filter paper and grids were air-dried. Samples were visualized using a JEOL JEM-2100Plus (JEOL) TEM and an accelerating voltage of 200 kV. Images were analysed and annotated using ImageJ (https://imagej.net/Fiji).

Phage DNA extraction

Nuclease-free H_2O was added to 200 µl of concentrated phage for a total final volume of 450 µl. This was then incubated at 37 °C for 1.5 h with 50 µl of 10×CutSmart Buffer (New England BioLabs) supplemented with 5 mM $CaCl_2$, 10 µl of DNase I (1 U μl^{-1}) (Thermo Scientific) and RNase A (10 mg ml^{-1}) (Thermo Scientific). Next, 20 µl of EDTA (final concentration 20 mM) and 1.3 µl of Proteinase K (20 mg ml^{-1}) (Qiagen) were added and incubated at 56 °C for 1.5 h. The Qiagen DNeasy Blood and Tissue Kit (Qiagen) was used to extract and purify the phage DNA. DNA was eluted in 20 µl of the kit's AE buffer. Phage DNA integrity was checked using agarose gel electrophoresis (1% agarose gel in 1×TAE buffer; 70 V, 90 min) against a GeneRuler 1 kb ladder (ThermoFisher Scientific).

Phage DNA sequencing, genome assembly and characterisation

Sequence data were generated on our in-house Illumina MiSeq platform. Extracted DNA was adjusted to a concentration of 0.2 ng/µL and treated using the Nextera XT DNA library preparation kit (Illumina) to produce fragments of approximately 500 bp. Fragmented and indexed samples were run on the sequencer using a Micro flow cell with the MiSeq Reagent Kit v2

(Illumina; 150 bp paired-end reads) following Illumina's recommended denaturation and loading procedures. Quality of raw sequence data was assessed using FastQC v0.11.9. Reads had a mean phred score above 30 and contained no adapter sequences, so data were not trimmed. Genomes were assembled using SPAdes v3.13.0 (default settings) [31], and visualized using Bandage v0.8.1 [32]. Contamination and completeness of genomes were determined using CheckV v0.8.1 [33]. Paired-end sequence reads were also mapped against assembled genomes using Geneious Prime. Genomes were screened for antimicrobial resistance genes using the Resistance Gene Identifier (v5.2.0) of the Comprehensive Antibiotic Resistance Database (v3.1.4) [34].

ViPTree v1.9 [35] was used to determine whether the phage genomes were closely related to previously described double-stranded DNA viruses. Initial analyses showed them to be closely related to *Klebsiella virus KP15* (genus *Slopekvirus*) [36]. Other *Slopekvirus* genomes were identified in GenBank and from the literature (Table 2) and included in a second ViPTree analysis. All publicly available genome sequences were also compared against the genomes of *Klebsiella* phage vB_KmiM-2Di and *Klebsiella* phage vB_KmiM-4Dii using pyani v0.2.11 (ANIm) [37].

Analysis of homing endonucleases (HENs) encoded within phage genomes

Genes in all genomes included in the initial ViPTree analysis (Table 2) were predicted and annotated using Prokka 1.14.6 [38] using the PHROG [39] dataset. Data were imported into R using Biostrings v2.58.0 [40] and predicted protein names searched for 'HNH|homing' to identify hypothetical HENs encoded with the phage genomes. The HEN sequences were exported in fasta format and imported into Geneious Prime. An MSA was created using Clustal Omega v1.2.3 (options selected: group sequences by similarity, evaluate full distance matrix, five refinement iterations). From the MSA, a neighbour-joining tree (Jukes–Cantor) was generated. The tree was uploaded to iToL v6.4.2 [41] for annotation. All data associated with the HEN analysis are available from figshare.

Comparative genome analyses

PhageClouds uses a gene-network approach to allow rapid searching of ~640000 phage genomes via a web interface, and includes metagenome-assembled genomes (MAGs) derived from several large-scale metagenome and virome studies [42]. Sequences of our initial set of GenBank genomes (Table 2) were searched against the PhageClouds database to identify slopekvirus genomes not included in our original analysis, and to determine whether such genomes had been detected in metagenome/virome studies. Results from PhageClouds searches were manually checked to identify a non-redundant set of genomes potentially representing slopekviruses. Quality and completeness of the genomes were determined using CheckV. ViPTree was used to confirm the newly identified genomes fell within the genus *Slopekvirus*.

For all genomes found to be of high quality or complete (n=24; Table 2), VIRIDIC was used to calculate the intergenomic similarities of the virus sequences [43]. The 24 genomes were Prokka-annotated as described above and included in a pangenome analysis (Roary v3.12.0, 95% identity [44]; . Using treeio v1.18.1 [45] and ggtree v3.2.1 [46] a phylogenetic tree was generated from the Roary-generated newick file (accessory_binary_genes.fa.newick), while the binary gene presence/absence file (gene_presence_absence.Rtab) was used to visualize the core and accessory genes identified in the slopekvirus pangenome. For each of the 24 genome sequences, nucleotide and amino acid sequences of core genes with ≥95% nucleotide identity and ≥70% coverage (based on comparisons of minimum, average and maximum gene group sizes determined from Roary outputs) were concatenated. These sequences were then used to generate MSAs with MUSCLE v3.8.1551 [47]. Maximum-likelihood trees were generated from the MSAs using PHYML v3.3.20180214 (BLOSUM62 for protein sequences, 100 bootstraps; Jukes–Cantor for nucleotide sequences, 100 bootstraps [48]), and visualized using iToL v6.4.2. The Prokka-annotated genomes and outputs from Roary are available from figshare.

RESULTS

Morphological and genomic characterization of phages isolated on *K. michiganensis*

Two phages (vB_KmiM-2Di and vB_KmiM-4Dii) had been isolated and purified on *K. michiganensis* strains PS_Koxy2 and PS_Koxy4, respectively, during ongoing studies focussed on finding alternatives to antibiotics for treating *Klebsiella*-associated infections [25]. In this study, genome sequence data and TEM images were generated for both phages. The genomes were found to be of high quality and free of contamination using CheckV (Table 2): the linear genome of vB_KmiM-2Di (148×coverage) was 99.34% complete, comprising 177200 nt and encoding 275 genes; the linear genome of vB_KmiM-4Dii (177×coverage) was 98.02% complete, comprising 174857 nt and encoding 271 genes. Mapping of the sequence reads against their respective genomes showed the even coverage of reads expected for circularly permuted genomes (not shown). We did not observe paired reads from clones across the ends, further indicating that our assembled genomes were incomplete. Neither phage-encoded antimicrobial resistance genes. Initial BLAST-based and ViPTree analyses suggested the genomes represented members of the genus *Slopekvirus* (*Duplodnaviria* › *Heunggongvirae* › *Uroviricota* › *Caudoviricetes* › *Straboviridae* › *Slopekvirus*). A ViPTree analysis, incorporating all slopekvirus genomes known to us at the time of analysis, confirmed this

Table 2. *Slopekvirus* genomes included in this study

Phage	Genome accession	Genome size (nt)	CDS*	CheckV			Reference
				Quality†	Completeness (%)	Contamination (%)	
Klebsiella phage vB_KmiM-2Di	MZ707156	177200	275	High-quality	99.34	0	This study
Klebsiella phage vB_KmiM-4Dii	MZ707157	174857	272	High-quality	98.02	0	This study
Klebsiella virus KP15	NC_014036	174436	269	High-quality	97.79	0	[49]
Klebsiella virus KP27	NC_020080	174413	271	High-quality	97.78	0	[49]
Klebsiella virus Matisse	NC_028750	176081	273	High-quality	98.73	0	[66]
Escherichia phage phT4A	NC_055712	171598	259	High-quality	96.20	0	[67]
Enterobacter virus Eap3	NC_041980	175814	275	High-quality	98.56	0	[68]
Klebsiella virus Miro	NC_041981	176055	274	High-quality	98.70	0	[69]
Klebsiella virus PMBT1	NC_042138	175206	274	High-quality	98.22	0	[70]
Klebsiella phage vB_KpM-KalD	LR881140	174351	268	Complete	100	0	[21]
Klebsiella phage vB_KpM-SoFaint	LR881141	175933	271	Complete	100	0	[21]
Klebsiella phage vB_KpM-Milk	LR881142	176734	271	Complete	100	0	[21]
Klebsiella phage vB_KoM-Liquor	LR881143	176734	270	Complete	100	0	[21]
Klebsiella phage vB_KoM-Pickle	LR881145	175221	273	High-quality	98.23	0	[21]
Klebsiella phage vB_KpM-Mild	LR881147	176856	271	Complete	100	0	[21]
Klebsiella phage vB_KoM-MeTiny	LR883651	175419	272	Complete	100	0	[21]
Klebsiella phage vB_KpnM_05F‡	LR746310	174231	272	Complete	100	0	[30]
Klebsiella phage KOX8‡	MN101221	131200	190	Medium-quality	73.55	0	Unpublished
Klebsiella phage KOX10‡	MN101223	168074	252	High-quality	94.22	0	Unpublished
Klebsiella phage KOX11‡	MN101224	59583	88	Low-quality	33.42	0	Unpublished
Klebsiella phage vB_KpnM_P-KP2‡	MT157285	172138	262	High-quality	96.51	0	[71]
Klebsiella phage vB_KpnM_M1‡	MW448170	176329	275	High-quality	98.85	0	Unpublished
SAMN00791912_a1_ct12238‡	–	174367	272	High-quality	97.76	0	[72]
uvig_221808‡	–	18523	17	Genome-fragment	10.78	0	[73]
uvig_279022‡	–	16907	33	Genome-fragment	9.48	0	[73]
uvig_279024‡	–	15914	37	Genome-fragment	8.93	0	[73]

Continued

Table 2. Continued

Phage	Genome accession	Genome size (nt)	CDS*	CheckV			Reference
				Quality†	Completeness (%)	Contamination (%)	
uvig_279033‡	–	12820	25	Genome-fragment	7.19	0	[73]
uvig_298784‡	–	175996	274	High-quality	98.67	0	[73]
uvig_335830‡	–	174400	272	High-quality	97.77	0	[73]
uvig_373971‡	–	16673	24	Genome-fragment	7.20	22.93	[73]
uvig_510342‡	–	18306	38	Genome-fragment	10.26	0	[73]
uvig_510345‡	–	15123	34	Genome-fragment	8.48	0	[73]
Zuo_2017_SRR5677802_NODE_1_length_175069_cov_4814.635012‡	–	175069	272	High-quality	100	0	[74]

*Annotated using the PHROG hmm database with Prokka.
†Complete, completeness assessed DTR (high-confidence); for all other annotations (i.e. high-quality, genome fragment) completeness was assessed by AAI (high-confidence). High-quality, >90% completeness; medium-quality, 50–90% completeness; low-quality, 0–50% completeness; genome-fragment, undetermined quality (no completeness estimate available) [33].
‡Detected using PhageClouds.

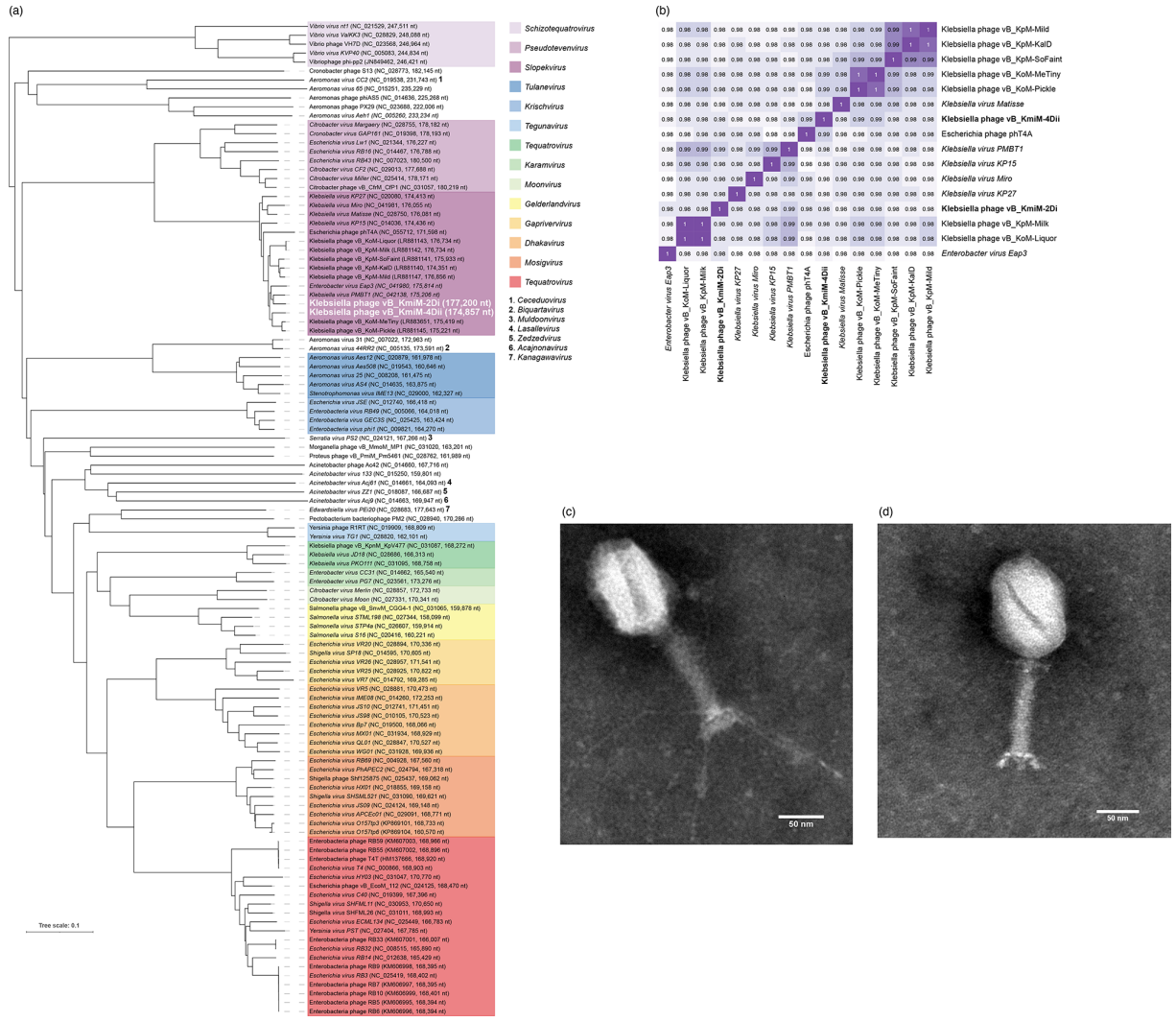

Fig. 1. Genomic analyses of slopekviruses *Klebsiella* phage vB_KmiM-2Di and *Klebsiella* phage vB_KmiM-4Dii and related viruses. (a) ViPTree-generated phylogenetic tree of slopekviruses and their closest relatives. (b) ANIm analysis of vB_KmiM-2Di and vB_KmiM-4Dii genomes with members of the genus *Slopekvirus*. (c) Phage vB_KmiM-2Di. (d) Phage vB_KmiM-4Dii. (c, d) Scale bar, 50 nm.

association (Fig. 1a). vB_KmiM-2Di clustered most closely with *Klebsiella virus PMBT1* and *Enterobacter virus Eap3*, whereas vB_KmiM-4Dii was most closely related to *Klebsiella* phage vB_KoM-MeTiny and *Klebsiella* phage vB_KoM-Pickle. ANIm analysis (Fig. 1b) showed vB_KmiM-2Di shared ≥99% ANI with *Klebsiella virus PMBT1*. vB_KmiM-4Dii shared ≥99% ANI with *Escherichia* phage phT4A, *Klebsiella* phage vB_KoM-MeTiny and *Klebsiella* phage vB_KoM-Pickle.

TEM images (Fig. 1c and d) showed both phages were myoviruses due to their contractile tails. Phage vB_KmiM-2Di (Fig. 1c) had a clearly visible base plate and long tail fibres. The capsid diameter was 84.1 nm (SD 3.8 nm), the tail and baseplate were 135.73 nm in length (SD 5.3 nm) and the phage had a total length of 251.6 nm (SD 11.5 nm). Phage vB_KmiM-4Dii (Fig. 1d) had a capsid diameter of 84.7 nm (SD 2 nm), its tail and baseplate were 130.9 nm in length (SD 1 nm) and the total length was 240.24 nm (SD 14.4 nm) (n=4 measurements for each phage). Three short tail fibres could be seen attached at the bottom of the base plate and short whiskers protruded from the collar under the capsid.

Identification of clinical and veterinary KoC isolates

rpoB gene sequence data were generated for clinical (n=59), and veterinary (n=49) isolates previously identified as *K. oxytoca* using MALDI-TOF MS and API 20E tests (Table 1). Sequence analysis of the *rpoB* gene can be used to differentiate between members of the KoC, *Klebsiella pneumoniae* complex and *Raoultella* spp. using a single primer set. Alternative methods, such as

differences in the bla_{OXY} gene, cannot be used for differentiation of the *K. pneumoniae* complex or other closely related organisms such as members of the genus *Raoultella* [10, 17, 18].

These data were compared against the 45 *rpoB* reference allele sequences (released 19 June 2020) available for download from the PubMLST *Klebsiella oxytoca*/*michiganensis*/*grimontii*/*pasteurii* typing database and *rpoB* gene sequences of the type strains of *Klebsiella* (including *Raoultella*) spp. (Fig. S1). This was done to confirm the identity of isolates as we (and others) have previously shown that phenotypic tests and MALDI-TOF are often inadequate for characterization of KoC isolates [10].

Analysis of the sequence data revealed *K. michiganensis* was the most prevalent species represented in both clinical and veterinary isolates (46% and 43%, respectively; 27/59 clinical, 21/49 veterinary; Fig. 2a), with the *rpoB* gene sequences of all isolates clustering in the clade with the *rpoB* gene sequence of *K. michiganensis* W14^T (Fig. S1). *K. oxytoca* was the second most prevalent species represented in the clinical isolates (40%; 24/59), but only represented 6% of veterinary isolates (3/49). *K. grimontii* was the second most-common bacterium among the veterinary isolates (24%; 12/49) and third most-prevalent species in the clinical isolates (12%; 7/59). *R. ornithinolytica* (12%; 6/49), *K. huaxiensis* (4%; 2/49) and *R. terrigena* (2%; 1/49) were only represented in veterinary isolates. *K. pneumoniae* represented 2% (1/59) of clinical isolates and 8% (4/49) of veterinary isolates.

The majority (73%; 36/49) of veterinary isolates were of bovine origin, isolated predominantly from milk-related samples (61%; 30/49) (Fig. 2b). While isolates of *K. grimontii* (8/36), *K. huaxiensis* (2/36), *K. oxytoca* (1/36), *K. pneumoniae* (2/36), *R. ornithinolytica* (4/36) and *R. terrigena* (1/36) had been recovered from milk samples, the majority (44%; 16/36) of the bovine isolates were *K. michiganensis* (Fig. 2c).

Host range analysis

vB_KmiM-2Di and vB_KmiM-4Dii were screened against clinical (*n*=59) and veterinary (*n*=49) isolates using an agar overlay method (Table 3). Phage activity against each strain was recorded as ++, complete lysis; +, hazy lysis; 0, no visible plaques. Depolymerase activity (*d*) was also recorded.

Both phages displayed a broad host range against clinical and veterinary isolates tested. vB_KmiM-2Di showed lytic activity against 66% (71/108) of tested strains (78%, 46/59 clinical and 51%, 25/49 veterinary strains, respectively). Whereas vB_KmiM-4Dii showed lytic activity against 84% (91/108) of strains (92%, 54/59 clinical and 76%, 37/49 veterinary strains, respectively). The two phages showed lytic activity against one or more strains of *K. michiganensis*, *K. oxytoca*, *K. grimontii*, *R. ornithinolytica*, *K. huaxiensis* and *R. terrigena*. Additionally, vB_KmiM-4Dii showed activity against *K. pneumoniae*.

The formation of haloes around plaques, indicating depolymerase activity, was observed for both phages against strains Ko13 and GFKo7. vB_KmiM-4Dii also displayed depolymerase activity against strain GFKo41. Neither vB_KmiM-2Di nor vB_KmiM-4Dii was able to infect 17% (18/108) of the strains tested, the majority (80%, 12/15) of which were of veterinary origin.

Analysis of HENs encoded within slopekvirus genomes

It has previously been suggested that differences in host range for members of the genus *Slopekvirus* may be due to homing endonucleases (HENs) encoded within their genomes. These HENs may act as regulators of DNA modification and provide resistance to host restriction enzymes [49]. Fifty-four putative HENs were encoded within 13/16 *Slopekvirus* genomes (Fig. 3). Several genomes encoded multiple putative HENs, with *Escherichia* phage phT4A (NC_055712) encoding nine. vB_KmiM-2Di encoded five HENs whereas vB_KmiM-4Dii encoded four. Of these, two had possible homologues in both genomes (Fig. S2). MZ707156_00002 shared homology with MZ707157_00002; 79.7% pairwise amino acid identity. MZ707156_00164 shared homology with MZ707157_00163; 100% pairwise amino acid identity.

Diversity of slopekviruses in Genank and metagenomic data

PhageClouds was used to identify slopekviruses not included in our initial scoping of viruses related to phages vB_KmiM-2Di and vB_KmiM-4Dii. Each PhageClouds search generated 33 hits. Table S1 provides example output for a search done using the genome sequence of phage vB_KmiM-2Di. Seventeen of the 33 hits represented sequences included in our initial phylogenetic analysis (Fig. 1a), six represented isolated phages not included in our initial list of viruses and 11 represented MAGs recovered from virome studies (Table 2). ViPTree analysis showed all 17 viruses were related to the genus *Slopekvirus* (data not shown). CheckV analysis showed four of the GenBank and four of the MAG sequences represented complete or high-quality phage genomes (Table 2).

Our initial analyses (Fig. 1, Fig. S3) showed that slopekvirus genomes shared high sequence similarity, with the HENs contributing to their diversity. Consequently, we used all 24 high-quality/complete slopekvirus genomes identified in this study (Table 2) in a pangenome analysis to determine whether core genes could be identified within the genus *Slopekvirus* (Fig. 4). It has been recommended that sequence and similarity coverage of proteins are set to >30% identity and 50% coverage, respectively, for genus-level phage-based pangenome studies [50]. However, for the genus *Slopekvirus* we found these criteria were too lax (not shown). Our Roary-based pangenome analysis run at 95% identity identified 155 core genes in a total pangenome of 425 genes

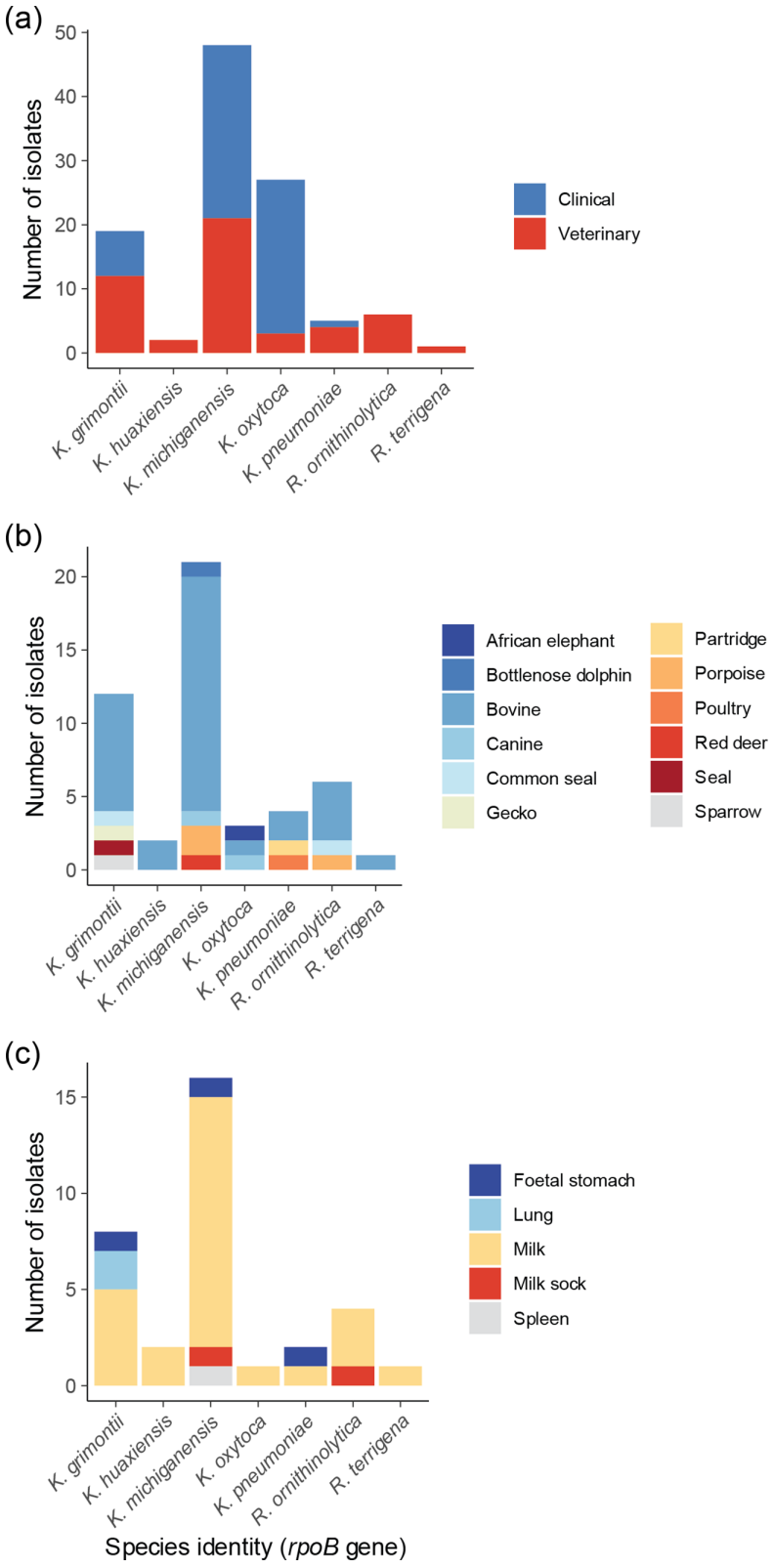

Fig. 2. Graphical representations of sources of clinical and veterinary isolates included in this study. (a) Distribution of clinical (*n*=59) and veterinary (*n*=49) isolates among the nine species of bacteria identified in this study. (b) Association of veterinary isolates (*n*=49) recovered from different animals with the nine species of bacteria identified in this study. (c) Association of bovine-associated isolates (*n*=34) with eight different species of bacteria identified in this study.

Table 3. Host range analysis of phage vB_KmiM-2di and phage vB_KmiM-4Dii

Lab ID*	*rpoB* gene-based identity	vB_KmiM-2Di	vB_KmiM-4Dii	Lab ID*	*rpoB* gene-based identity	vB_KmiM-2Di	vB_KmiM-4Dii
Ko1	*K. oxytoca*	+†	+	GFKo2	*K. grimontii*	0	+
Ko2	*K. oxytoca*	+	+	GFKo3	*K. grimontii*	0	+
Ko3	*K. michiganensis*	+	++	GFKo4	*R. ornithinolytica*	++	++
Ko4	*K. oxytoca*	+	++	GFKo5	*K. grimontii*	+	+
Ko5	*K. michiganensis*	+	+	GFKo6	*K. oxytoca*	0	0
Ko6	*K. oxytoca*	+	+	GFKo7	*K. michiganensis*	+,d	++,d
Ko7	*K. oxytoca*	0	+	GFKo8	*K. michiganensis*	0	+
Ko8	*K. grimontii*	+	+	GFKo9	*R. ornithinolytica*	0	0
Ko9	*K. grimontii*	+	++	GFKo10	*K. michiganensis*	0	+
Ko10	*K. michiganensis*	0	+	GFKo11	*K. huaxiensis*	+	+
Ko11	*K. oxytoca*	++	++	GFKo12	*K. grimontii*	0	0
Ko12	*K. oxytoca*	+	++	GFKo13	*K. michiganensis*	0	0
Ko13	*K. michiganensis*	+,d	+,d	GFKo14	*K. michiganensis*	+	+
Ko14	*K. michiganensis*	0	+	GFKo15	*K. michiganensis*	+	++
Ko15	*K. oxytoca*	+	++	GFKo16	*K. oxytoca*	++	++
Ko16	*K. grimontii*	++	++	GFKo17	*R. ornithinolytica*	0	0
Ko17	*K. oxytoca*	+	++	GFKo18	*K. michiganensis*	+	+
Ko18	*K. michiganensis*	++	++	GFKo19	*K. michiganensis*	+	+
Ko19	*K. oxytoca*	++	++	GFKo20	*K. grimontii*	++	++
Ko20	*K. oxytoca*	+	0	GFKo21	*K. grimontii*	++	++
Ko21	*K. michiganensis*	+	+	GFKo22	*K. grimontii*	+	+
Ko22	*K. michiganensis*	+	+	GFKo24	*R. ornithinolytica*	0	+
Ko23	*K. michiganensis*	+	+	GFKo25	*R. ornithinolytica*	0	+
Ko24	*K. michiganensis*	+	+	GFKo26	*K. grimontii*	+	+
Ko25	*K. oxytoca*	0	+	GFKo27	*K. grimontii*	+	+
Ko26	*K. oxytoca*	++	++	GFKo28	*K. michiganensis*	+	++
Ko27	*K. grimontii*	+	++	GFKo29	*K. michiganensis*	++	+
Ko28	*K. michiganensis*	+	+	GFKo30	*K. michiganensis*	0	0

Continued

Table 3. Continued

Lab ID*	*rpoB* gene-based identity	vB_KmiM-2Di	vB_KmiM-4Dii	Lab ID*	*rpoB* gene-based identity	vB_KmiM-2Di	vB_KmiM-4Dii
Ko29	*K. michiganensis*	+	+	GFKo31	*K. michiganensis*	+	+
Ko30	*K. grimontii*	0	+	GFKo32	*K. michiganensis*	0	+
Ko31	*K. michiganensis*	+	++	GFKo33	*K. grimontii*	0	0
Ko32	*K. michiganensis*	+	+	GFKo34	*K. grimontii*	+	+
Ko33	*K. michiganensis*	0	+	GFKo35	*K. michiganensis*	+	+
Ko34	*K. pneumoniae*	0	+	GFKo36	*K. michiganensis*	+	++
Ko35	*K. michiganensis*	+	0	GFKo37	*K. michiganensis*	+	+
Ko36	*K. michiganensis*	0	+	GFKo38	*K. michiganensis*	0	++
Ko37	*K. oxytoca*	+	+	GFKo39	*K. grimontii*	++	++
Ko38	*K. oxytoca*	+	+	GFKo40	*K. michiganensis*	+	+
Ko39	*K. michiganensis*	+	+	GFKo41	*K. michiganensis*	0	+,d
Ko40	*K. oxytoca*	0	0	GFKo42	*K. pneumoniae*	0	+
Ko41	*K. michiganensis*	0	0	GFKo43	*R. ornithinolytica*	0	0
Ko42	*K. grimontii*	+	+	GFKo44	*K. michiganensis*	0	0
Ko43	*K. michiganensis*	+	+	GFKo45	*K. michiganensis*	+	+
Ko44	*K. oxytoca*	+	+	GFKo46	*K. pneumoniae*	0	+
Ko45	*K. oxytoca*	+	+	GFKo47	*K. oxytoca*	0	0
Ko46	*K. michiganensis*	0	0	GFKo48	*K. pneumoniae*	0	+
Ko47	*K. oxytoca*	+	+	GFKo49	*K. pneumoniae*	0	0
Ko48	*K. oxytoca*	+	+	GFKo50	*K. huaxiensis*	0	0
Ko49	*K. michiganensis*	0	+	GFKo52	*R. terrigena*	++	++
Ko50	*K. michiganensis*	+	+				
Ko51	*K. grimontii*	+	+				
Ko52	*K. michiganensis*	+	+				
Ko53	*K. oxytoca*	+	+				
Ko54	*K. oxytoca*	0	+				
Ko55	*K. oxytoca*	+	+				

Continued

Table 3. Continued

Lab ID*	*rpoB* gene-based identity	vB_KmiM-2Di	vB_KmiM-4Dii	Lab ID*	*rpoB* gene-based identity	vB_KmiM-2Di	vB_KmiM-4Dii
Ko56	*K. michiganensis*	+	+				
Ko57	*K. oxytoca*	+	+				
Ko58	*K. michiganensis*	+	+				
Ko59	*K. michiganensis*	+	+				

*Ko prefix, clinical isolate; GFKo prefix, veterinary isolate.
† ++,† ++, complete lysis; +, hazy lysis; 0, no visible plaques; *d*, depolymerase activity.

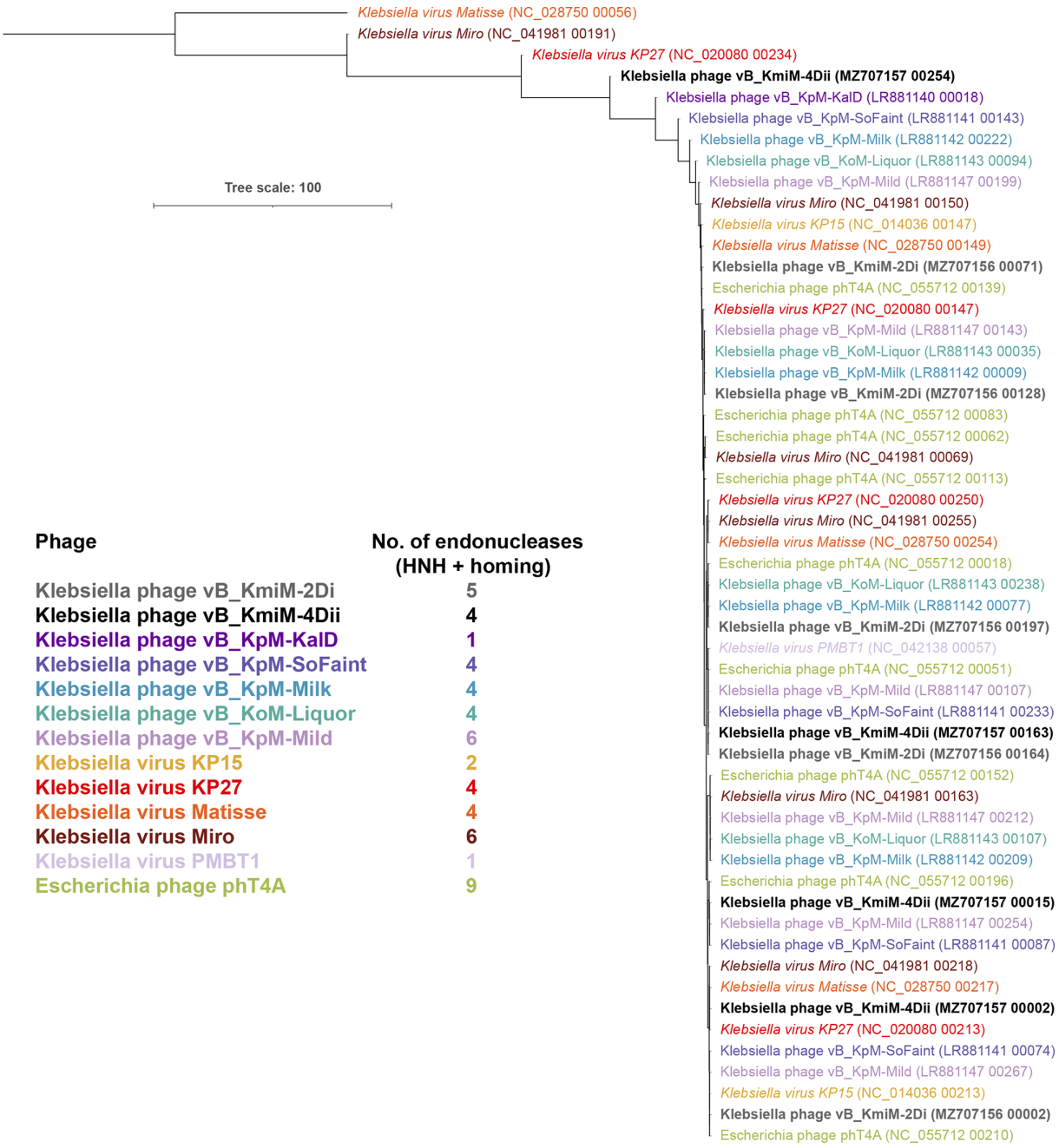

Fig. 3. Phylogenetic analysis (neighbour-joining tree) and summation of endonucleases encoded in slopekviruses. In the tree each gene product is coloured based on the phage that encodes it. Scale bar, average number of amino acid substitutions per position. The alignment, distance matrix and newick file associated with the analysis are available from figshare.

(Fig. 4a and b, Table S2). The pangenome was open, as the plot showing the total number of genes is not asymptotic (Fig. 4c). Filtering the core genes based on comparisons of minimum, maximum and average group nucleotide coverage (Table S2) identified 148 core genes had ≥95% identity and ≥70% coverage. Given that the 24 genomes encoded a mean of 272 genes (+/- 5 genes) each, these core genes represented 54% (148/272) of the total genome content of the slopekviruses. The majority (61/148, 41%) of these genes were predicted to encode hypothetical proteins, with baseplate wedge subunit proteins (6/148, 4%), proteins of unknown function (4/148, 3%), tail tube (3/148, 2%) and RIIB lysis inhibitor, major head, head scaffolding, clamp loader of DNA polymerase, baseplate hub and 5′–3′ deoxyribonucleotidase proteins (2/148, 1% each) making the greatest contribution to the core genes (Fig. S4). As expected, the hypothetical HENs contributed to the accessory genes.

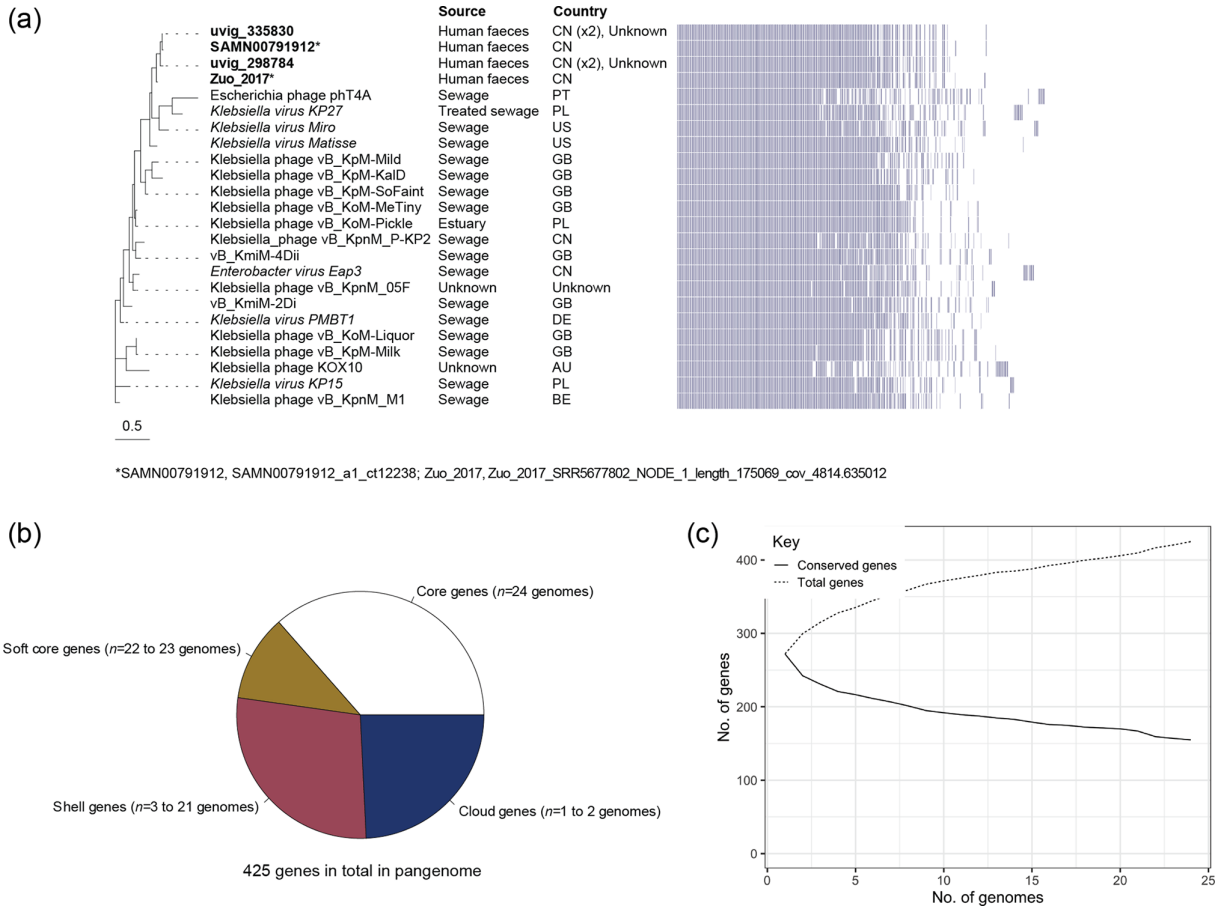

Fig. 4. Visualization of pangenome data for 24 slopekvirus genomes. (a) Roary matrix of the 425 genes representing the total slopekvirus pangenome. The dendrogram was generated from the Roary output file 'accessory_binary_genes.fa'. Isolation source and country are shown for each phage. AU, Australia; BE, Belgium; CN, China; GB, Great Britain; DE, Denmark; PL, Poland; PT, Portugal; US, United States of America. (×2), phage reported to have been detected in two different metagenomic datasets from the same country. (b) Pie chart showing the contribution of the 425 genes to the pangenome: core genes, found in all 24 genomes analysed; soft core genes, found in 22–23 genomes analysed; shell genes, found in between 3 and 21 genomes analysed; cloud genes, found in 1–2 genomes analysed. (c) Plots showing how the number of total genes (dotted line) and conserved genes (dashed line) changes as more genomes are added to the slopekvirus pangenome.

If two phage genome sequences, tested reciprocally, are more than 95% identical at the nucleotide level over their full genome length, they are assigned to the same species [50]. We, therefore, sought to determine the species diversity within the genus *Slopekvirus*. VIRIDIC analysis of the 24 genome sequences included in our pangenome analysis suggested there were eight species represented within the genus (Table S3), though bidirectional hierarchical clustering of these data showed no obvious clustering of the species (Fig. 5a). An MSA of the concatenated nucleotide sequences for the 148 core genes (available from figshare) showed the total alignment of 99, 156 nt shared a minimum identity of 96.41% across all genomes (Table S4). Maximum-likelihood phylogenetic analysis (Fig. 5b) of these concatenated sequences did not support species separation, nor did clustering of the accessory gene (binary) data in the pangenome analysis (Fig. 4a). An analysis based on the concatenated amino acid sequences of the 148 core genes showed similar results (Fig. S5).

DISCUSSION

In this study we characterized the morphology and genomes of two phages (vB_KmiM-2Di and vB_KmiM-4Dii) we had isolated on two strains (PS_Koxy2 and PS_Koxy4, respectively) of multidrug-resistant *K. michiganensis* [10, 25]. Both phages were found to belong to the genus *Slopekvirus* and had the myovirus-type morphology consistent with other known slopekviruses (Fig. 1). The host ranges of these phages were determined on a collection of clinical (n=59) and veterinary (n=49) bacteria that had been originally identified as belonging to the KoC (Table 1). Both phages were found to have broad host ranges (Table 3). This feature of the slopekviruses is thought to be due to the number of HENs encoded within their genomes (Fig. 3). Searches among ~640000 phage genomes, using PhageClouds, allowed us to identify slopekviruses within metagenome and virome datasets.

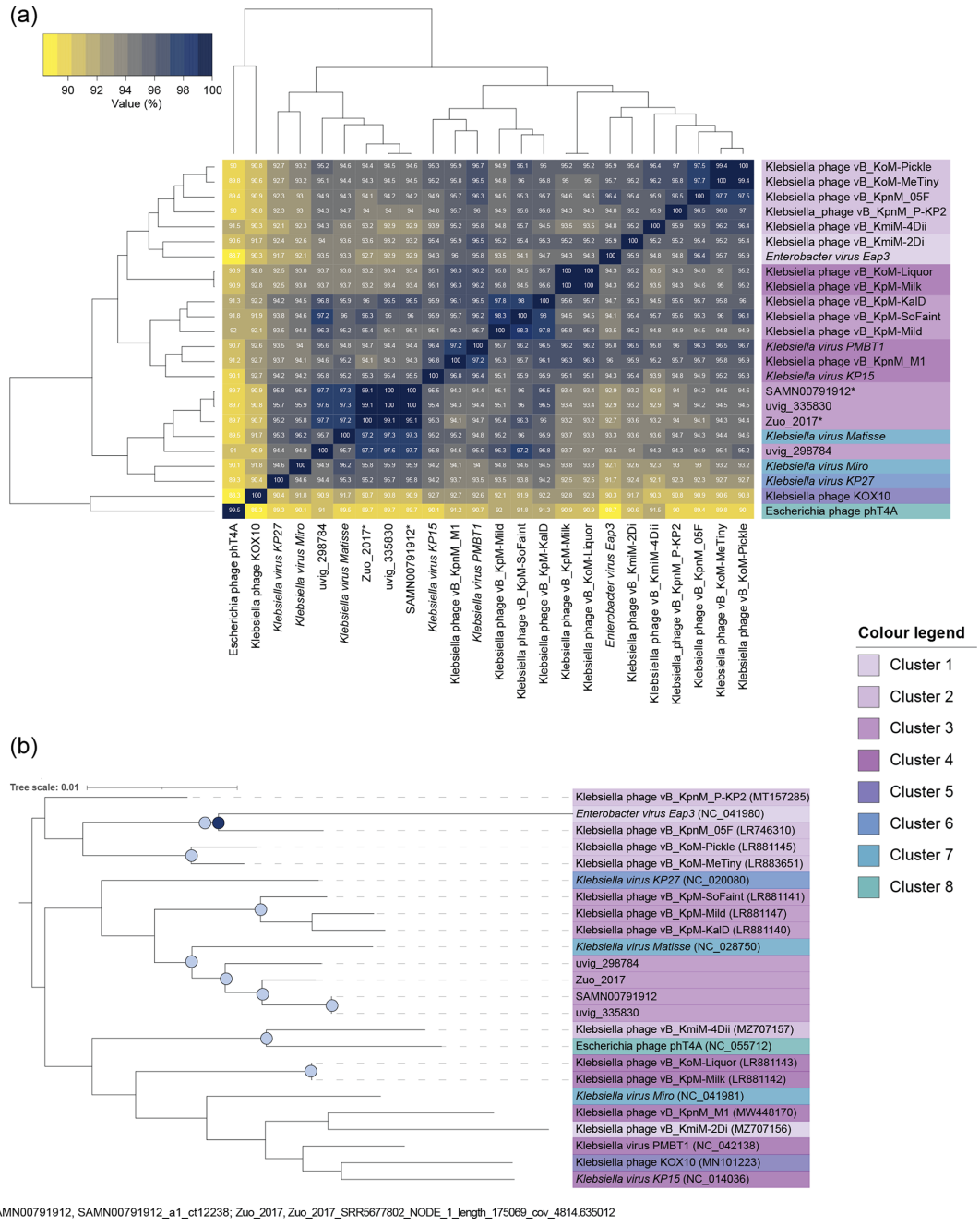

Fig. 5. Determination of whether distinct species are represented with the genus *Slopekvirus*. (a) Bidirectional clustering heatmap visualizing VIRIDIC-generated similarity matrix for the 24 slopekvirus genomes. (b) Phylogenetic tree (maximum-likelihood, Jukes–Cantor neighbour-joining initial tree) generated by concatenating the nucleotide sequences of the 148 core genes from the pangenome that had ≥95% identity and ≥70% coverage across the 24 genomes. Dark-blue circle: 100% bootstrap support by maximum likelihood; light-blue circles, 100% bootstrap support by neighbour-joining only. The tree was rooted at the mid-point. (a, b) Colouring of the data corresponds to the eight species clusters predicted by VIRIDIC.

We also characterized the genomes of our phages and those of their closest relatives, undertaking a pangenome analysis to better understand the genomic diversity of slopekviruses.

Identification of members of the KoC from clinical and veterinary sources

Despite our clinical and veterinary isolates being presumptively identified as *K. oxytoca* by MALDI-TOF MS and API 20E profiling, *K. oxytoca* only represented 25% (27/108) of the isolates identified by *rpoB* gene sequence analysis. While we have previously highlighted that phenotypic assays and MALDI-TOF MS frequently do not allow accurate identification of clinical isolates of the

KoC [10, 51], this is the first time we have encountered issues identifying veterinary KoC isolates. Veterinary MALDI-TOF MS databases should be updated to include relevant reference spectra to allow differentiation of KoC species [1].

The *rpoB* gene-based sequence analysis identified *K. michiganensis* as the most common isolate from both clinical and veterinary samples (27/59 and 21/49, respectively). This supports our previous findings with respect to *K. michiganensis* being more clinically relevant than *K. oxytoca* [51]. Gómez *et al.* also used *rpoB* gene sequence analysis to identify commensal, community-acquired and neonatal intensive care unit *Klebsiella* spp. isolates. Routine biochemical testing identified 21 *K. oxytoca*; however, *rpoB* gene analysis identified them as *K. michiganensis* (16/21), *K. grimontii* (5/21) and *K. pneumoniae* (1/21) [18]. Taken together with our work, these findings reinforce the clinical relevance of *K. michiganensis* and its historic underappreciation as a human pathogen due to misidentification.

In recent years, *K. michiganensis* has been isolated from a diverse range of animals: farm animals (cows, poultry and pigs) [13, 52], companion animals (cat, dog and horses) [12, 52] and other animals including hedgehogs, guineapigs, mice, fruit bats, turtles and invertebrates [13, 52–55]. This is from using tools capable of discriminating between members of the KoC, but further work is needed to assess the true clinical and epidemiological significance of *K. michiganensis* in animals. Furthermore, many of these studies are retrospective, using accurate but time- and resource-intensive methods that would not be practical for rapid clinical or veterinary diagnosis. It is notable that 16/21 of our *K. michiganensis* veterinary isolates had been recovered from bovine milk samples (Table 1; Fig. 2c). Whether these isolates encode genetic determinants specific to mastitis, as observed for an acquired *lac* operon in bovine-associated *K. pneumoniae* [56], will be determined when we analyse genome sequence data for them in the future, with results to be reported elsewhere.

K. grimontii has previously been isolated from animals in Germany (cattle/milk, rabbit, pig, sheep, dog, pig, tortoise, hedgehog, roe deer) [13], and during a longitudinal study undertaken in Pavia, Italy (sheep, horse, fly, cattle, pig, cat, duck, turtle, dog, cockroach, wasp, chicken) [52]. We found *K. grimontii* in seal (*n*=2), bovine (*n*=8), gecko (*n*=1) and sparrow (*n*=1) samples collected in the UK. *K. grimontii* appears to be the second most-common KoC species of veterinary relevance.

K. huaxiensis was originally described based on one isolate recovered from human urine in China [16]. Since then, the bacterium has been isolated from cow and human faeces in Italy [1], and from cows (*n*=4), water (*n*=2), a horse (*n*=1) and hospital carriage (*n*=1) [52]. In this study we identified two isolates (GFKo11, GFKo50) from bovine milk collected in Scotland. As for other members of the KoC, more work is needed to determine the wider relevance of *K. huaxiensis* to veterinary infections.

Host range determination for phages vB_KmiM-2Di and vB_KmiM-4Dii

We found that phages vB_KmiM-2Di and vB_KmiM-4Dii both exhibited broad host ranges against several different species of bacteria (Table 3). This includes isolates of *K. pneumoniae*, *K. michiganensis*, *K. oxytoca*, *K. grimontii*, *R. ornithinolytica*, *K. huaxiensis* and *R. terrigena*. Ordinarily, phage host range is narrow, sometimes down to the strain level. However broad-host-range phages are reported in the literature [57–59] and phages with extended host ranges have been identified within the genus *Slopekvirus* [21, 36]. For example, phage vB_KoM-MeTiny, which is genetically similar to vB_KmiM-2Di and vB_KmiM-4Dii (95.2% and 96.2% respectively; VIRIDIC), is reported to form plaques on *K. michiganensis*, *K. oxytoca*, *K. pneumoniae*, *K. varicola* and *K. quasipneumoniae*. These findings suggest bacteriophage belonging to the genus *Slopekvirus* are useful as potential therapeutics owing to their broad host ranges and lack of antimicrobial resistance genes [21, 36], in agreement with others who have worked with slopekviruses [21, 30].

Despite a high level of sequence identity between vB_KmiM-2Di and vB_KmiM-4Dii (95.25%; VIRIDIC) differences in host range were observed between the two phages. Maciejewska *et al.* previously characterized two slopekviruses (vB_KpnM_KP15 and vB_KpnM_KP27) that exhibit broad lysis against *Klebsiella* spp. [49]. They also noted differences in the host ranges of the two phages, despite a high-level of DNA identity (94.2%; our VIRIDIC analysis). The authors suggested the discord in host range may be due to the presence of two HENs encoded within the genome of vB_KpnM_KP27, both of which are absent from vB_KpnM_KP15. HENs are site-specific DNA endonucleases that function as mobile genetic elements by catalysing a double-strand break at specific DNA target sites in a recipient genome that lacks the endonuclease. The double-strand break is then repaired by homologous recombination using the allele containing the HEN as the template. The result is the incorporation of the HEN into the cleavage site. Maciejewska *et al.* hypothesized that HENs encoded within vB_KpnM_KP27 (YP_007348875.1 and YP_007348891.1) may act as regulators of DNA modification by splicing events near DNA modification genes located in their close vicinity. Such events may result in protection against host restriction enzymes and subsequently modulate host range. This could occur through alteration of genes involving nucleotide metabolism, DNA methylation or DNA restriction. We identified a homologue of YP_007348875.1 (86% pairwise amino acid identity with MZ707157_00254) and its associated flanking region in the genome of vB_KmiM-4Dii that is absent from vB_KmiM-2Di (Fig. S6). The flanking regions invB_KmiM-4Dii are identical to those found in phage KP27 described in Maciejewska *et al.* This includes the upstream RNA ligase and downstream DprA-like DNA recombination-mediator protein. The second HEN identified in KP27 (YP_007348891.1) was absent from both vB_KmiM-2Di and vB_KmiM-4Dii. The presence of YP_007348875.1 in the genome of vB_KmiM-4Dii may help explain its slightly broader host range compared to vB_KmiM-2Di.

We also conducted a search for other HENs encoded in other slopekvirus genomes. Despite a high degree of genetic conservation across the genus (Fig. 4, Fig. 5), there was a high level of divergence with respect to the number and type of HENs encoded across the 16 genomes and between vB_KmiM-2Di and vB_KmiM-4Dii. We found that vB_KmiM-2Di encodes five hypothetical HENs, whereas vB_KmiM-4Dii encodes four. Of note was a HEN identified in the genomes of both vB_KmiM-2Di and vB_KmiM-4Dii (locus tags MZ707156_00164 and MZ707157_00163, respectively; Prokka-annotated data available from figshare). This HEN was found immediately upstream of a CDS encoding a putative SbcC-like subunit of a predicted palindrome-specific endonuclease (Fig. S7). This may again be a case of a HEN splicing event near a gene involved in DNA modification. SbcC and SbcD (also present in the phage genome) are nucleases that have been shown to be involved in the degradation of palindromic DNA structures and associated hairpin structures [60].

In addition to exploring the possibility of HENs affecting the host range of the two phages, we also sought to identify differences in tail fibre proteins, which are often determinants of host-cell specificity. We examined the genomic region encoding five tail fibre proteins as identified by Maciejewska *et al.* [49] (Fig. S8a). Our analysis found the first three of these proteins share a high level of amino acid identity between homologues in vB_KmiM-2Di and vB_KmiM-4Dii (>90% pairwise identity). However, the fourth protein, encoding the tail fibre (predicted as an L-shaped tail fibre protein by HHPred), shares only 76.80% pairwise amino acid identity. There was a considerable lack of conservation at the C-terminal of this protein, which is considered to be the region responsible for host-receptor recognition [61]. Additionally, the final protein of this region did not appear to be homologous between the phages (13.3% pairwise amino acid identity). As determined by HHPred, the predicted tail fibre adhesin of vB_KmiM-2Di shows structural similarity to protein gp38 from *Salmonella* phage S16, which is involved in host-cell receptor recognition [62]. However, the tail fibre assembly protein of vB_KmiM-4Dii is a chaperone involved in correct folding of the tail fibre and likely trimerization [63]. Amino acid alignments for tail fibre proteins are presented inFig. S8b. These results may further explain the differences in host range observed between vB_KmiM-2Di and vB_KmiM-4Dii. These findings lend support to the theory that HENs, in combination with differences in tail fibre proteins, may influence host range and contribute to the differences in host lysis observed between vB_KmiM-2Di and vB_KmiM-4Dii.

It has been noted by others [21, 64] that phages with highly similar sequences may have altered lytic spectra due to selection pressures applied by different hosts used for propagation. In the current study, *K. michiganensis* strains PS_Koxy2 and PS_Koxy4 were used for propagation of vB_KmiM-2Di and vB_KmiM-4Dii, respectively. These two strains have been extensively characterized both genomically and phenotypically and have been shown to share >99.9% ANI and the same multi-locus sequence type [10]. Therefore, we conclude that any differences observed with respect to host range are unlikely to be a consequence of the host used for propagation.

Genetic diversity of slopekviruses: implications for phage taxonomy

Our Roary-based pangenome analysis of 24 high-quality/complete slopekvirus genomes showed the genus *Slopekvirus* comprises phages with highly conserved genomes, with the core genome (determined using ≥95% identity and ≥70% coverage) representing over half of the total genome content of the genus (Fig. 4). The recommended cut-off criteria (>30% identity, 50% coverage [50]) for genus-level phage pangenome analysis was found to be inappropriate for use with the genus *Slopekvirus*. Results from our analysis (along with other unpublished work from our laboratory) suggest wider-ranging studies are required to determine appropriate recommended identity and coverage cut-off values for use in phage-based pangenome studies at any taxonomic level.

The use of VIRIDIC (and other ANI tools) to assign phage that share ≥95% reciprocal sequence identity at the nucleotide level over their full genome length to the same species is also questioned [50]. Our VIRIDIC analysis of 24 high-quality/complete slopekvirus genomes suggested our dataset represented eight different species (Fig. 5). However, phylogenetic analysis based on the nucleotide sequences of 148 core genes (encoding 99, 156 nt) did not support separation of the genus into eight species. ANI analyses must be supplemented with phylogenetic analyses to provide robust evidence to support 'species' designations within phage genera, as recommended for bacterial and archaeal taxonomy [65].

Funding information
P.S. was in receipt of an IBMS Research Grant (project title "Isolation of lytic bacteriophages active against antibiotic-resistant Klebsiella pneumoniae"). Imperial Health Charity is thanked for contributing to registration fees for the Professional Doctorate studies of P.S. L.H. was funded by UK Med-Bio (Medical Research Council grant number MR/L01632X/1). SRUC Veterinary Services receive funding from the Scottish Government as part of its Veterinary Advisory Services programme.

Acknowledgements
This work used computing resources funded by the Research Contingency Fund of the Department of Biosciences, Nottingham Trent University (NTU). T.S.Z. completed this work as part of an MRes degree at NTU. We thank Andrew Millard (University of Leicester) for providing us with the Prokka formatted PHROG dataset for annotation of phage genomes.

Author contributions
Conceptualization: P.S., L.H., D.N. Data curation: all authors. Formal analysis: T.S.Z., P.S., A.L.M., G.F., L.H., D.N. Funding acquisition: P.S., A.L.M., L.H. Investigation: T.S.Z., P.S., A.L.M., L.H., D.N. Methodology: A.L.M., L.H., D.N. Resources: G.F., A.L.M., L.H., D.N. Supervision: A.L.M., L.H., D.N. Visualization: T.S.Z., L.H., D.N. Writing – original draft: T.S.Z., L.H., D.N. Writing – reviewing and editing: all authors.

Conflicts of interest
The author(s) declare that there are no conflicts of interest.

Ethical statement
The study of anonymised clinical isolates provided by the Nottingham University Hospitals NHS Trust (NUH) Pathogen Bank was approved by NUH Research and Innovation (19MI001).

References

1. Merla C, Rodrigues C, Passet V, Corbella M, Thorpe HA, *et al*. Description of *Klebsiella spallanzanii* sp. nov. and of *Klebsiella pasteurii* sp. nov. *Front Microbiol* 2019;10:2360.
2. Granier SA, Leflon-Guibout V, Goldstein FW, Nicolas-Chanoine MH. New *Klebsiella oxytoca* beta-lactamase genes bla(OXY-3) and bla(OXY-4) and a third genetic group of *K. oxytoca* based on bla(OXY-3). *Antimicrob Agents Chemother* 2003;47:2922–2928.
3. Fevre C, Jbel M, Passet V, Weill F-X, Grimont PAD, *et al*. Six groups of the OXY beta-lactamase evolved over millions of years in *Klebsiella oxytoca*. *Antimicrob Agents Chemother* 2005;49:3453–3462.
4. Youn Y, Lee SW, Cho H-H, Park S, Chung H-S, *et al*. Antibiotics-associated hemorrhagic colitis caused by *Klebsiella oxytoca*: two case reports. *Pediatr Gastroenterol Hepatol Nutr* 2018;21:141–146.
5. Abadi RM. Emerging carbapenemase *Klebsiella oxytoca* with multidrug resistance implicated in urinary tract infection. *Biomed Biotechnol Res J* 2020;4:148–151.
6. England PH. MSSA and Gram-negative Bacteraemia and CDI: Annual Report; 2020 Dec. https://www.gov.uk/government/statistics/mrsa-mssa-and-e-coli-bacteraemia-and-c-difficile-infection-annual-epidemiological-commentary [accessed 18 January 2021].
7. Moradigaravand D, Martin V, Peacock SJ, Parkhill J. Population structure of multidrug resistant *Klebsiella oxytoca* within hospitals across the UK and Ireland identifies sharing of virulence and resistance genes with *K. pneumoniae*. *Genome Biol Evol* 2017;9:574–587.
8. Yang J, Long H, Hu Y, Feng Y, McNally A, *et al*. *Klebsiella oxytoca* complex: update on taxonomy, antimicrobial resistance, and virulence. *Clin Microbiol Rev* 2022;35:e0000621.
9. Ellington MJ, Davies F, Jauneikaite E, Hopkins KL, Turton JF, *et al*. A multispecies cluster of GES-5 carbapenemase-producing Enterobacterales linked by a geographically disseminated plasmid. *Clin Infect Dis* 2020;71:2553–2560.
10. Shibu P, McCuaig F, McCartney AL, Kujawska M, Hall LJ, *et al*. Improved molecular characterization of the *Klebsiella oxytoca* complex reveals the prevalence of the kleboxymycin biosynthetic gene cluster. *Microb Genom* 2021;7:000592.
11. Lee D, Oh JY, Sum S, Park HM. Prevalence and antimicrobial resistance of *Klebsiella* species isolated from clinically ill companion animals. *J Vet Sci* 2021;22:1–13.
12. Loncaric I, Cabal Rosel A, Szostak MP, Licka T, Allerberger F, *et al*. Broad-spectrum cephalosporin-resistant *Klebsiella* spp. isolated from diseased orses in Austria. *Animals* 2020;10:332.
13. Klaper K, Hammerl JA, Rau J, Pfeifer Y, Werner G. Genome-based analysis of *Klebsiella* spp. isolates from animals and food products in Germany, 2013-2017. *Pathogens* 2021;10:573.
14. Bridel S, Watts SC, Judd LM, Harshegyi T, Passet V, *et al*. *Klebsiella* MALDI TypeR: a web-based tool for *Klebsiella* identification based on MALDI-TOF mass spectrometry. *Res Microbiol* 2021;172:103835.
15. Boye K, Hansen DS. Sequencing of 16S rDNA of *Klebsiella*: taxonomic relations within the genus and to other *Enterobacteriaceae*. *Int J Med Microbiol* 2003;292:495–503.
16. Hu Y, Wei L, Feng Y, Xie Y, Zong Z. *Klebsiella huaxiensis* sp. nov., recovered from human urine. *Int J Syst Evol Microbiol* 2019;69:333–336.
17. Passet V, Brisse S. Description of *Klebsiella grimontii* sp. nov. *Int J Syst Evol Microbiol* 2018;68:377–381.
18. Gómez M, Valverde A, Del Campo R, Rodríguez JM, Maldonado-Barragán A. Phenotypic and molecular characterization of commensal, community-acquired and Nosocomial *Klebsiella* spp. *Microorganisms* 2021;9:2344.
19. Herridge WP, Shibu P, O'Shea J, Brook TC, Hoyles L. Bacteriophages of *Klebsiella* spp., their diversity and potential therapeutic uses. *J Med Microbiol* 2020;69:176–194.
20. Li P, Zhang Y, Yan F, Zhou X. Characteristics of a bacteriophage, vB_Kox_ZX8, isolated from clinical *Klebsiella oxytoca* and its therapeutic effect on mice bacteremia. *Front Microbiol* 2021;12:763136.
21. Townsend EM, Kelly L, Gannon L, Muscatt G, Dunstan R, *et al*. Isolation and characterization of *Klebsiella* phages for phage therapy. *Phage* 2021;2:26–42.
22. Ku H, Kabwe M, Chan HT, Stanton C, Petrovski S, *et al*. Novel *Drexlerviridae* bacteriophage KMI8 with specific lytic activity against *Klebsiella michiganensis* and its biofilms. *PLoS One* 2021;16:e0257102.
23. Zamani I, Bouzari M, Emtiazi G, Ghasemi SM, Chang H-I. Molecular investigation of two novel bacteriophages of a facultative methylotroph, *Raoultella ornithinolytica*: first report of *Raoultella* phages. *Arch Virol* 2019;164:2015–2022.
24. Fofanov MV, Morozova VV, Kozlova YN, Tikunov AY, Babkin IV, *et al*. *Raoultella* bacteriophage RP180, a new member of the genus *Kagunavirus*, subfamily *Guernseyvirinae*. *Arch Virol* 2019;164:2637–2640.
25. Shibu P. *Investigations of carbapenem-resistant Klebsiella species and associated clinical considerations. PhD thesis*. , University of Westminster 2019.
26. Herzog KAT, Schneditz G, Leitner E, Feierl G, Hoffmann KM, *et al*. Genotypes of *Klebsiella oxytoca* isolates from patients with nosocomial pneumonia are distinct from those of isolates from patients with antibiotic-associated hemorrhagic colitis. *J Clin Microbiol* 2014;52:1607–1616.
27. Jolley KA, Bray JE, Maiden MCJ. Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications. *Wellcome Open Res* 2018;3:124.
28. Ma Y, Wu X, Li S, Tang L, Chen M, *et al*. Proposal for reunification of the genus *Raoultella* with the genus *Klebsiella* and reclassification of *Raoultella electrica* as *Klebsiella electrica* comb. nov. *Res Microbiol* 2021;172:103851.
29. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res* 2019;47:W256–W259.
30. Haines MEK, Hodges FE, Nale JY, Mahony J, van Sinderen D, *et al*. Analysis of selection methods to develop novel phage therapy cocktails against antimicrobial resistant clinical isolates of bacteriaClinical Isolates of Bacteria. *Front Microbiol* 2021;12:564.
31. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, *et al*. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 2012;19:455–477.
32. Wick RR, Schultz MB, Zobel J, Holt KE. Bandage: interactive visualization of *de novo* genome assemblies. *Bioinformatics* 2015;31:3350–3352.
33. Nayfach S, Camargo AP, Schulz F, Eloe-Fadrosh E, Roux S, *et al*. CheckV assesses the quality and completeness of metagenome-assembled viral genomes. *Nat Biotechnol* 2021;39:578–585.
34. Alcock BP, Raphenya AR, Lau TTY, Tsang KK, Bouchard M, *et al*. CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database. *Nucleic Acids Res* 2020;48:D517–D525.
35. Nishimura Y, Yoshida T, Kuronishi M, Uehara H, Ogata H, *et al*. ViPTree: the viral proteomic tree server. *Bioinformatics* 2017;33:2379–2380.
36. Kęsik-Szeloch A, Drulis-Kawa Z, Weber-Dąbrowska B, Kassner J, Majkowska-Skrobek G, *et al*. Characterising the biology of novel lytic bacteriophages infecting multidrug resistant *Klebsiella pneumoniae*. *Virol J* 2013;10:1–12.

37. Pritchard L, Glover RH, Humphris S, Elphinstone JG, Toth IK. Genomics and taxonomy in diagnostics for food security: soft-rotting enterobacterial plant pathogens. *Anal Methods* 2016;8:12–24.
38. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 2014;30:2068–2069.
39. Terzian P, Olo Ndela E, Galiez C, Lossouarn J, Pérez Bucio RE, *et al*. PHROG: families of prokaryotic virus proteins clustered using remote homology. *NAR Genom Bioinform* 2021;3:lqab067.
40. Pagès H, Aboyoun P, Gentleman R, DebRoy S. Biostrings: efficient manipulation of biological strings. R package version 2.64.0; 2022.
41. Letunic I, Bork P. Interactive tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res* 2021;49:W293–W296.
42. Rangel-Pineros G, Millard A, Michniewski S, Scanlan D, Sirén K, *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.
43. Moraru C, Varsani A, Kropinski AM. VIRIDIC-A Novel Tool to Calculate the Intergenomic Similarities of Prokaryote-Infecting Viruses. *Viruses* 2020;12:E1268.
44. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, *et al*. Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics* 2015;31:3691–3693.
45. Wang L-G, Lam TT-Y, Xu S, Dai Z, Zhou L, *et al*. Treeio: an R package for phylogenetic tree input and output with richly annotated and associated data. *Mol Biol Evol* 2020;37:599–603.
46. Yu G, Smith DK, Zhu H, Guan Y, ggtree L. Ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol Evol* 2017;8:28–36.
47. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 2004;32:1792–1797.
48. Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, *et al*. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* 2010;59:307–321.
49. Maciejewska B, Roszniowski B, Espaillat A, Kęsik-Szeloch A, Majkowska-Skrobek G, *et al*.*Klebsiella* phages representing a novel clade of viruses with an unknown DNA modification and biotechnologically interesting enzymes. *Appl Microbiol Biotechnol* 2017;101:673–684.
50. Turner D, Kropinski AM, Adriaenssens EM. A roadmap for genome-based phage taxonomy. *Viruses* 2021;13:506.
51. Chen Y, Brook TC, Soe CZ, O'Neill I, Alcon-Giner C, *et al*. Preterm infants harbour diverse *Klebsiella* populations, including atypical species that encode and produce an array of antimicrobial resistance- and virulence-associated factors. *Microb Genom* 2020;6.
52. Thorpe H, Booton R, Kallonen T, Gibbon MJ, Couto N, *et al*. One health or three? transmission modelling of *Klebsiella* isolates reveals ecological barriers to transmission between humans, animals and the environment. *Microbiology*; bioRxiv. DOI: 10.1101/2021.08.05.455249
53. Cai Z, Guo Q, Yao Z, Zheng W, Xie J, *et al*. Comparative genomics of *Klebsiella michiganensis* BD177 and related members of *Klebsiella* sp. reveal the symbiotic relationship with *Bactrocera dorsalis*. *BMC Genet* 2020;21:Suppl .
54. McDougall FK, Wyres KL, Judd LM, Boardman WSJ, Holt KE, *et al*. Novel strains of *Klebsiella africana* and *Klebsiella pneumoniae* in Australian fruit bats (*Pteropus poliocephalus*). *Res Microbiol* 2021;172:103879.
55. Oliveira RA, Ng KM, Correia MB, Cabral V, Shi H, *et al*.*Klebsiella michiganensis* transmission enhances resistance to *Enterobacteriaceae* gut invasion by nutrition competition. *Nat Microbiol* 2020;5:630–641.
56. Holt KE, Wertheim H, Zadoks RN, Baker S, Whitehouse CA, *et al*. Genomic analysis of diversity, population structure, virulence, and antimicrobial resistance in *Klebsiella pneumoniae*, an urgent threat to public health. *Proc Natl Acad Sci U S A* 2015;112:E3574–81.
57. Pan Y-J, Lin T-L, Chen C-C, Tsai Y-T, Cheng Y-H, *et al*. *Klebsiella* phage ΦK64-1 encodes multiple depolymerases for multiple host capsular types. *J Virol* 2017;91:e02457-16.
58. Verma V, Harjai K, Chhibber S. Characterization of a T7-like lytic bacteriophage of *Klebsiella pneumoniae* B5055: a potential therapeutic agent. *Curr Microbiol* 2009;59:274–281.
59. Wu L-T, Chang S-Y, Yen M-R, Yang T-C, Tseng Y-H. Characterization of extended-host-range pseudo-T-even bacteriophage Kpp95 isolated on *Klebsiella pneumoniae*. *Appl Environ Microbiol* 2007;73:2532–2540.
60. Wendel BM, Cole JM, Courcelle CT, Courcelle J. SbcC-SbcD and ExoI process convergent forks to complete chromosome replication. *Proc Natl Acad Sci U S A* 2018;115:349–354.
61. Garcia-Doval C, van Raaij MJ. Structure of the receptor-binding carboxy-terminal domain of bacteriophage T7 tail fibers. *Proc Natl Acad Sci USA* 2012;109:9390–9395.
62. Dunne M, Denyes JM, Arndt H, Loessner MJ, Leiman PG, *et al*. *Salmonella* phage S16 tail fiber adhesin features a rare polyglycine rich domain for host recognition. *Structure* 2018;26:1573–1582.
63. North OI, Sakai K, Yamashita E, Nakagawa A, Iwazaki T, *et al*. Phage tail fibre assembly proteins employ a modular structure to drive the correct folding of diverse fibres. *Nat Microbiol* 2012;4:1645–1653.
64. Jensen EC, Schrader HS, Rieland B, Thompson TL, Lee KW, *et al*. Prevalence of broad-host-range lytic bacteriophages of *Sphaerotilus natans*, *Escherichia coli*, and *Pseudomonas aeruginosa*. *Appl Environ Microbiol* 1998;64:575–580.
65. Chun J, Oren A, Ventosa A, Christensen H, Arahal DR, *et al*. Proposed minimal standards for the use of genome data for the taxonomy of prokaryotes. *Int J Syst Evol Microbiol* 2018;68:461–466.
66. Provasek VE, Lessor LE, Cahill JL, Rasche ES, Kuty Everett GF. Complete genome sequence of carbapenemase-producing *Klebsiella pneumoniae* myophage Matisse. *Genome Announc* 2015;3:e01136-15.
67. Pereira C, Moreirinha C, Teles L, Rocha RJM, Calado R, *et al*. Application of phage therapy during bivalve depuration improves *Escherichia coli* decontamination. *Food Microbiol* 2017;61:102–112.
68. Zhao J, Zhang Z, Tian C, Chen X, Hu L, *et al*. Characterizing the biology of lytic bacteriophage vB_EaeM_ϕEap-3 infecting multidrug-resistant *Enterobacter aerogenes*. *Front Microbiol* 2019;10:420.
69. Mijalis EM, Lessor LE, Cahill JL, Rasche ES, Kuty Everett GF. Complete genome sequence of *Klebsiella pneumoniae* carbapenemase-producing *K. pneumoniae* myophage miro. *Genome Announc* 2015;3:e01137-15.
70. Koberg S, Brinks E, Fiedler G, Hüsing C, Cho G-S, *et al*. Genome sequence of *Klebsiella pneumoniae* bacteriophage PMBT1 isolated from raw sewage. *Genome Announc* 2017;5:e00914-16.
71. Wang Z, Cai R, Wang G, Guo Z, Liu X, *et al*. Combination therapy of phage vB_KpnM_P-KP2 and gentamicin combats acute pneumonia caused by K47 serotype *Klebsiella pneumoniae*. *Front Microbiol* 2021;12:674068.
72. 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 U S A* 2021;118:e2023202118.
73. Camarillo-Guerrero LF, Almeida A, Rangel-Pineros G, Finn RD, Lawley TD. Massive expansion of human gut bacteriophage diversity. *Cell* 2021;184:1098–1109.
74. Gregory AC, Zablocki O, Zayed AA, Howell A, Bolduc B, *et al*. The gut virome database reveals age-dependent patterns of virome diversity in the human gut. *Cell Host Microbe* 2020;28:724–740.

Edited by: G. H. Thomas