K. pneumoniae ST258 genomic variability and bacteriophage susceptibility

Carola Venturini[1#,*], Nouri Ben Zakour[1*], Bethany Bowring[1], Sandra Morales[2], Robert Cole[2], Zsuzsanna Kovach[2], Steven Branston[2], Emma Kettle[3], Nicholas Thomson[4,5], Jonathan Iredell[1#]

[1]Centre for Infectious Diseases and Microbiology, The Westmead Institute for Medical Research, The University of Sydney and Westmead Hospital, Sydney, Australia;

[2]AmpliPhi Australia Pty Ltd, Sydney, Australia;

[3]Westmead Research Hub Electron Microscope Core Facility, The Westmead Institute for Medical Research, The University of Sydney, Sydney, Australia;

[4]Parasites and Microbes, Wellcome Trust Sanger Centre, Cambridge, United Kingdom;

[5]The London School of Hygiene and Tropical Medicine, London, United Kingdom.

[*]These authors contributed equally to this work.

Short title: *K. pneumoniae* ST258 and its bacteriophages

[#]**Corresponding authors:** Carola Venturini, Centre for Infectious Diseases and Microbiology, The Westmead Institute for Medical Research, Westmead NSW 2145, Australia. Tel.: 0061 2 86273415. Email: carola.venturini@sydney.edu.au; Jonathan R Iredell Centre for Infectious Diseases and Microbiology, The Westmead Institute for Medical Research, Westmead NSW 2145, Australia. Tel.: 0061 2 86273411. Email: jonathan.iredell@sydney.edu.au

Abstract

Multidrug resistant carbapenemase-producing *Klebsiella pneumoniae* capable of causing severe disease in humans is classified as an urgent threat by health agencies worldwide. Bacteriophages are being actively explored as potential therapeutics against these multidrug resistant pathogens. We are currently developing bacteriophage therapy against carbepenem-resistant *K. pneumoniae* belonging to the genetically diverse, globally disseminated clonal group CG258. In an effort to define a robust experimental approach for effective selection of lytic viruses for therapy, we have fully characterized the bacterial genomes of 18 target strains, tested them against novel lytic bacteriophages, and generated phage-susceptibility profiles. The genomes of *K. pneumoniae* isolates carrying bla_{NDM} and bla_{KPC} were sequenced and isolates belonging to CG258 were selected for susceptibility testing using a panel of lytic bacteriophages (n=65). The local *K. pneumoniae* CG258 population was dominated by isolates belonging to sequence type ST258 clade 1 (86%). The primary differences between ST258 genomes were variations in the capsular locus (*cps*) and in prophage content. We showed that CG258-specific lytic phages primarily target the capsule, and that successful infection is blocked in many, post-adsorption, by immunity conferred by existing prophages. Five bacteriophages specifically active against *K. pneumoniae* ST258 clade 1 (n=5) belonging to the Caudovirales order were selected for further characterization. Our findings show that effective control of *K. pneumoniae* CG258 with phage will require mixes of diverse lytic viruses targeting all relevant *cps* variants and allowing for variable prophage content. These insights will facilitate identification and selection of therapeutic phage candidates against this serious pathogen.

Importance

Bacteriophages are natural agents that exclusively and selectively kill bacteria and have the potential to be useful in the treatment of multidrug resistant infections. *K. pneumoniae* CG258 is a main agent of life-threatening sepsis that is often resistant to last-line antibiotics. Our work highlights some key requirements for developing bacteriophage preparations targeting this pathogen. By defining the genomic profile of our clinical *K. pneumoniae* CG258 population and matching it with bacteriophage susceptibility patterns, we found that bacteriophage ability to lyse each strain correlates well with *K. pneumoniae* CG258 structural subtypes (capsule variants). This indicates that preparation of bacteriophage therapeutics targeting this pathogen should aim at including phages against each bacterial capsular subtype. This necessitates a detailed understanding of the diversity of circulating isolates in different geographical areas in order to make rational therapeutic choices.

Introduction

Klebsiella pneumoniae is an important ubiquitous Gram-negative species capable of causing disease in both humans and animals (1). The rise in recent decades of *K. pneumoniae* that are multi-drug resistant (MDR), including to last-line antibiotics such as carbapenems, has resulted in the classification of this species as an urgent threat to human health by health agencies worldwide and its recognition as an important antimicrobial resistance reservoir (2,3). Carbapenemase-producing (CP) MDR strains, carrying the bla_{KPC} and bla_{NDM} genes, can be asymptomatic residents of the human gut and a major cause of serious nosocomial infections worldwide associated with high morbidity and mortality (4-7).

The genetically diverse clonal group CG258, comprising sequence types ST258, ST11, ST512 and a few other SNP variants, is largely responsible for the global dissemination of MDR CP-*K. pneumoniae* (6,7). Population studies looking at the genomes of CG258 have shown that diversification within this clonal group is linked to a series of large-scale genomic rearrangements and an apparent high frequency of recombination, some of which result in switching or variation in the capsule polysaccharide-encoding (*cps*) locus (6-8). On the basis of this variation, the ST258 group can be subdivided into two separate lineages clade 1 and clade 2. ST258 clade 2 strains have been the main cause of disease outbreaks worldwide (7,9), while clade 1 uniquely predominated a recent Australian outbreak (9).

Alternative or adjuvant therapies to antibiotics against MDR pathogens are urgently needed (2). Naturally occurring lytic bacterial viruses (bacteriophages, phages) were recognised as effective therapeutic agents in the first decades of the 20th century, but were little valued by Western medicine during the antibiotic era (10). The rise in multidrug resistance, however, has renewed the interest in their potential for both decontamination and eradication of pathogens refractory to antibiotics. Lytic bacteriophages against problematic bacterial species can be readily isolated, but medical applicability is hindered by limited understanding of key

issues such as optimal clinical protocols, penetration, and resistance, as well as disappointing clinical trials using phage that have been associated with inconsistent protocols and poor targeting (10-12). Bacteriophages capable of lysing *K. pneumoniae*, including MDR strains have been described (13-15), with complete genomes for more than 80 full double-stranded (ds) DNA phages available in NCBI databases to date. However, no effective therapeutic product has yet reached the bedside.

We are currently exploring bacteriophage therapy against extended-spectrum-β-lactamase (ESBL) producing *Enterobacteriaceae* isolated in Australia from humans with the aim of defining a robust experimental protocol for the rapid design of effective targeted phage preparations. Here, we present the bacteriophage susceptibility profiles of *K. pneumoniae* ST258 isolated in Australia and show their correlation with genomic variation within this clonal group. We also report the full genome sequence of five *K. pneumoniae* ST258-specific lytic phages (AmPh_EK29, AmPh_EK52, AmPh_EK80, JIPh_Kp122, JIPh_Kp127) isolated from wastewater in Australia.

Methods

Bacterial isolates

In this study, we have fully characterized the bacterial genomes of ESBL *K. pneumoniae* CG258 strains from Australia (n=18, with n=16 CP-*K. pneumoniae*) and tested the infectivity of novel bacteriophages (n=65) selected from our existing libraries or isolated *de novo* from local environmental sources. All MDR *K. pneumoniae* isolates containing the carbapenem resistance genes commonly associated with CG258, bla_{NDM} or bla_{KPC} (16), in our extensive clinical collection were selected as potential target isolates for this study (Table 1).

Table 1. *K. pneumoniae* isolates characterized in this study.

Name (JIE)*	Patient Identifier	State	Collection†	Carbapenemase encoding gene‡	Antibiotic resistance phenotype**	Reference
2487	1	Vic	2012	none	AMK AMC AMP ATM CAZ CRO CIP SXT GEN TZPi TIM TOB TMP	(17)
2709	2	Vic	2012 (June 06)	*bla*$_{KPC}$	AMK AMC AMP ATM FEP CAZ CRO CIP SXT MEM MXF TZP TIM TOB TMP	(16)
2713	3	Vic	2012	*bla*$_{NDM}$	AMK AMC AMP ATM FEP CAZ CRO CIP SXT GEN MEMi TZP TIM TOB TMP	(17)
2733	4	NSW	2012	*bla*$_{KPC}$	AMK AMC AMP ATM FEP CAZ CRO CIP SXT GENi MEM TZP TIM TOB TMP	(16)
2740	2	Vic	2012 (June 21)	none	AMK AMC AMP ATM FEP CAZ CRO CIP SXT TZPi TIM TOB TMP	(16)
2771	5	NSW	2012	*bla*$_{KPC}$	AMK AMC AMP ATM CAZ CRO CIP MEM TZP TIM TOB TMPi	(16)
2783	6	NSW	2010	*bla*$_{KPC}$	AMK AMC AMP ATM FEP CAZ CRO CIP CST SXT MEM TZP TIM TOB TMP	(16)
2793	7	WA	2012	*bla*$_{KPC}$	AMK AMC AMP ATM CAZ CRO CIP SXT MEM TZP TIM TOB TMP	(16)
3095	8	Vic	2012	*bla*$_{KPC}$	AMK AMC AMP ATM FEP CAZ CRO CIP GENi MEM TZP TIM TOB TMPi	(16)
4005	9	Vic	2014 (Jan 09)	*bla*$_{KPC}$	AMK AMC AMP ATM FEP CAZ CRO CIP GENi MEM TZP TIM TOB	(16)
4019	9	Vic	2014 (Jan 31)	*bla*$_{KPC}$	AMK AMC AMP ATM FEP CAZ CRO CIP SXT MEM TZP TIM TOB TMP	(16)
4020	10	Vic	2014	*bla*$_{KPC}$	AMK AMC AMP ATM FEP CAZ CRO CIP GENi MEMi TZP TIM TOB	(16)
4046	11	Vic	2014	*bla*$_{KPC}$	AMKi AMC AMP ATM FEP CAZ CRO CIP MEM TZP TIM TOB	(16)

continues next page

Table 1 continued

Name (JIE)*	Patient Identifier	State	Collection†	Carbapenemase encoding gene‡	Antibiotic resistance phenotype**	Reference
4203	12	Vic	2014	*bla*$_{KPC}$	AMK AMC AMP ATM FEP CAZ CRO CIP MEM TZP TIM TOB TMP	(16)
4282	13	Vic	2014	*bla*$_{KPC}$	AMC AMP ATM FEP CAZ CRO CIP SXT GEN MEMi TZP TIM TOB TMP	(16)
4455	14	NSW	2015	*bla*$_{KPC}$	AMK AMC AMP ATM FEP CAZ CRO CIP SXT MEM TZP TIM TOB TMP	(16)
4626	16	NSW	2015	*bla*$_{KPC}$	AMKi AMC AMP ATM FEP CAZ CRO CIP CST SXT GEN MEM TZP TIM TOB TMP	(16)
4660	17	NSW	2015	*bla*$_{KPC}$	AMKi AMC AMP ATM FEP CAZ CRO CIP CST SXT GEN MEM TZP TIM TOB TMP	(16)

*from JIE_G1046886_Iredell collection; †in brackets, month and day specified for isolates collected from the same patient; ‡determined by diagnostic PCR screening at collection facility (ICPMR, Westmead Hospital, Westmead, NSW, Australia); **determined by BD Phoenix™ (Becton Dickinson, Wokingham, Berkshire, UK) screening at collection facility. Cut-off values in accordance with the EUCAST system. i, intermediate. For isolate 2793, FEP susceptibility was not determined.

112

113 **CP-*K. pneumoniae* phenotypes**

114 *Biofilm production*

115 Biofilm formation by growing bacteria in polypropylene microtitre plates was estimated by
116 crystal violet staining of adherent cells, following the protocol of O'Toole and Kolter (18)
117 with minor modifications. Briefly, overnight bacterial cultures in lysogeny broth (LB; Oxoid,
118 Basingstoke, UK) adjusted to OD_{600} 0.4 were added (0.1 mL) to microtiter plate (Corning
119 Life Sciences, Corning, NY, USA) wells and grown overnight in a static incubator at 37°C.
120 Wells were carefully washed twice with RO water before addition of 0.1% crystal violet
121 (Sigma-Aldrich, MO, USA) (225 µL) and incubation at room temperature. Plates were gently
122 washed four times with RO water and dried at room temperature for at least 2 h. For

quantitation, 200 µL of 100% ethanol was added to each well and left for 10-15 min. An aliquot (125 µL) of the solubilized solution was then transferred to a new flat bottom microtiter dish and absorbance at 540 nm was measured in a Spectromax Vmax microplate reader (Biomolecular Devices, San Jose, CA, USA). Negative (LB only) and positive (ATCC 27853 *P. aeruginosa*, a strong biofilm producer) controls were included on all plates. Experiments were performed in triplicate.

Polysaccharide capsule production

Total capsule production in *K. pneumoniae* CG258 was quantified according to previously described methods (19,20). Briefly, overnight bacterial cultures in Mueller-Hinton broth (Oxoid, Basingstoke, UK) were mixed with 1% Zwittergent 3-14 detergent (Millipore, Billerica, MA, USA) in 100 mM citric acid (pH 2.0) and incubated for 30 min at 50°C with occasional mixing. After pelleting the bacteria, 300 µL of supernatant was mixed with absolute ethanol to a final concentration of 80% and left on ice for 30 min to allow for capsule precipitation. After centrifugation, the precipitates were allowed to dry and then resuspended in 100 µL of DNase-free water (Lonza, Rockland, ME, USA) and kept at 4°C overnight. Capsule quantitation was assayed by measuring uronic acid content on ethanol-precipitated culture supernatants by addition of 1.2 mL of 12.5 mM tetraborate (Sigma-Aldrich, St. Louis, MO, USA) in concentrated sulphuric acid (Sigma-Aldrich, St. Louis, MO, USA), and detection (absorbance at 520 nm) using 0.15% m-hydroxydiphenyl (Sigma-Aldrich, St. Louis, MO, USA) in 0.5% NaOH (Amresco, Solon, OH, USA). Sodium hydroxide added to the tetraborate/sulphuric acid solution was used as the baseline for quantification. Capsule quantification was performed in triplicate for each bacterial strain.

148 *Lipopolysaccharide profiles*

149 Lipopolysaccharide (LPS) profiles of *K. pneumoniae* CG258 strains were analysed using
150 sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) subsequent to
151 proteinase K digestion of the proteins in whole cell lysates (21). Overnight bacterial cultures
152 (100 µl) were pelleted and washed twice in saline. The final pellet was resuspended in 35 µl
153 sterile saline. Pellets were treated with 20 µl 4x SDS reducing buffer (0.0625 M Tris-HCl, pH
154 8.8, 10% glycerol, 2% SDS, 5% 2-b-mercaptoethanol, 0.0125% bromophenol blue) at 100°C
155 for 10 min. Proteins were digested by addition of 15 µl 20 mg/ml proteinase K and incubation
156 at 60°C for 1 h. Sample preparations were run on a 15% separating gel with a 4% stacking
157 gel under reducing conditions using Tris-glycine running buffer (mini-PROTEAN system,
158 Bio-Rad Laboratories, Hercules, CA, USA). Silver staining using sodium thiosulphate
159 sensitisation and silver nitrate was performed according to established methods (22).

160

161 **Sequencing and analysis of bacterial genomes**

162 The genomes of selected MDR *K. pneumoniae* isolates were sequenced by Illumina NextSeq
163 (paired-end; 2 x 150 bp). Bacterial DNA extraction was performed using the DNeasy Blood
164 and Tissue DNA isolation kit (Qiagen, Hilden, Germany) to obtain high purity ($OD_{260/280}$ 1.8-
165 2.0; $OD_{260/230}$ 1.8) preparations for sequencing. DNA libraries for whole genome sequencing
166 (WGS) were prepared using the Nextera XT kit and sequencing was performed at the
167 Australian Genome Research Facility (AGRF, Melbourne, Australia). *De novo* assembly of
168 sequencing reads and simulated reads of NCBI reference genomes were performed as
169 previously described (23), using our WGS analysis workflow based on publicly available
170 tools including SPAdes 3.9.0 (24); Nullarbor 1.20 (25); Kleborate 0.2.0 (26) to confirm
171 identity (*in silico* MLST), virulence and antibiotic resistance genotypes. A maximum-
172 likelihood recombination-free phylogenetic tree was computed using RAxML 8.2.4 (27) and

Gubbins 2.2.0 (28), using a reference-based core genome alignment as an input. The publicly available genome sequences of five representative CG258 strains were also added for comparative purposes: AUSMDU00008079, used as mapping reference (CP022691); HS11286 (CP003200); NJST258-1 (CP006923); Kb140 (AQROD00000000); and VA360 (ANGI00000000). The pangenome was determined using Roary version 3.11.0 (29) and used to classify regions of differences across the strain dataset, based on their contiguity and functional categories of the genes encoded. Kleborate 0.2.0 (26) was used to perform capsule typing, O antigen (LPS) serotyping and siderophore typing. Plasmid replicon identification and typing was performed using PlasmidFinder and pMLST implemented in BAP (30). Prophage-associated contigs were annotated using PHASTER (31). Further *cps* locus comparative analysis was performed using EasyFig (32) and Geneious v9.1 (https://www.geneious.com). Gap closure between separate contigs in the capsular locus was achieved by PCR amplification and Sanger sequencing of purified linkage amplicons (AGRF, Melbourne, Australia).

bla$_{KPC}$ *genomic context*

To confirm plasmid content and genomic context of the *bla*$_{KPC}$ gene in target ST258 (n=16) isolates, we performed Pulse Field Gel Electrophoresis (PFGE) on S1 nuclease (Promega, Madison, WI, USA) digested DNA, as before (33,34), and Southern hybridization with *bla*$_{KPC}$ and *rep* IncFII$_K$ DIG-labelled probes (16) prepared using published primers (35,36) and the PCR DIG Probe Synthesis Kit (Roche, Mannheim, Germany) following manufacturer's instructions. Images were obtained on a ChemiDoc™ MP System (Bio-Rad Laboratories, Richmond, CA, USA).

197 **_De novo_ isolation of CP-_K. pneumoniae_ specific bacteriophages**

198 Bacteriophages against target *K. pneumoniae* CG258 were isolated from sewage and
199 wastewater samples collected in the Greater Sydney District (Sydney, NSW, Australia).
200 Specimens were clarified by centrifugation and filtration through a 0.22 µm filter. Aliquots of
201 environmental filtrates were then incubated with target *K. pneumoniae* isolates overnight.
202 Bacteriophages were selected from single plaques in double-layer agar assays and purified
203 through three rounds of plating (37). High-titre stocks were prepared by propagating phage
204 over several double-layer plates washed in SM buffer (50 mM Tris-HCl, 8 mM $MgSO_4$, 100
205 mM NaCl, pH 7.4) and filtered through a 0.22 µm filter. The concentration of phage-forming
206 units (PFU) per ml was determined by spotting 10 µl of ten-fold serial dilutions onto a
207 double-layer of the target bacteria (37). High-titre ($\geq 10^9$ PFU/mL) phage stocks were stored
208 at 4°C.

209

210 **Bacteriophage host range**

211 The identified CG258 *K. pneumoniae* strains were tested against phages (n=65) selected from
212 our extensive library or isolated *de novo* against one of the target isolates. Bacteriophage host
213 range was determined by measuring the efficiency of plating (EOP) for each phage-bacteria
214 combination. Ten-fold serial dilutions (10 µl) were spotted onto a double-layer of the target
215 bacteria and compared to the original isolation host (37). *Escherichia coli*, *Enterococcus*
216 *faecium*, *Pseudomonas aeruginosa*, *Staphylococcus aureus* and *Staphylococcus*
217 *pseudintermedius* were used as cross-species controls. In order to further test bacteriophage
218 specificity, we also blind tested the host ranges of three phages with unique specificities
219 (AmPh_EK52, AmPh_EK38 and JIPh_Kp122) on a set of CP-*K. pneumoniae* isolates from
220 Europe (n=48) to determine their predictive diagnostic value linked to *K. pneumoniae*
221 sequence or capsule type. Bacteriophage activity against each strain was scored as: 1. ‘full

222 activity' for presence of clear plaques at highest dilution; 2. 'poor activity' for presence of
223 turbid plaques, or isolated bacterial colonies within clearings, or EOP three or more $\log_{10}$
224 lower than that of the original host; 3. 'partial activity' for evidence of clearing in bacterial
225 lawn, but absence of distinct plaques; 4. 'negative' for very faint, difficult to observe,
226 clearing or absence of any visible plaques or clearing zones in the bacterial lawn.
227
228 **Bacteriophage characterization**
229 *Genome sequencing*
230 Bacteriophages (n=5) that specifically lysed *K. pneumoniae* ST258 clade 1 with different host
231 range profiles were selected for further characterization as potential therapeutic candidates.
232 Bacteriophage DNA was extracted using the Wizard DNA Clean-Up System (Promega,
233 Madison, WI, USA) and used for whole genome sequencing (WGS) (Nextera XT kit;
234 Illumina NextSeq, paired-end, 2 x 150 bp). Briefly, total DNA concentration was quantified
235 using Quant-it PicoGreen dsDNA Assay Kit (Invitrogen, Carlsbad, CA, USA) and 1 ng/µl of
236 DNA was used to prepare DNA libraries using the Nextera XT Library Preparation Kit and
237 Nextera XT v2 Indexes (Illumina, San Diego, CA, USA). Multiplexed libraries were
238 sequenced using paired end 150 bp chemistry on the NextSeq 500 NCS v2.0 (Illumina). Error
239 rates were calculated using PhiX Sequencing Control v3 for each run. De-multiplexing and
240 FastQC generation was performed with default settings using BaseSpace (Illumina).
241 Bacteriophage genomes were assembled using our in-house genomic pipeline and annotated
242 using RAST-tk (38). The absence of lysogeny modules, virulence and resistance determinants
243 was determined using our WGS analysis workflow (as for bacterial genomes) and PHASTER
244 (31). Genome comparisons with best database (GenBank, NCBI) matches were obtained
245 using EasyFig (32). PFGE of intact viral particles was performed to confirm relative size
246 (Chef Mapper System, Bio-Rad Laboratories, Hercules, CA, USA).

Imaging of bacteriophages

Bacteriophage preparations were dialysed against 0.1 M ammonium acetate in dialysis cassettes with a 10,000 membrane molecular weight cut-off (Pierce Biotechnology, Rockford, IL, USA), negatively stained with 2% uranyl acetate and visualised using transmission electron microscopy (TEM) (37). TEM was conducted at the Westmead Electron Microscopy Facility (Westmead, Australia) on a Philips CM120 BioTWIN transmission electron microscope at 100kV. Images were recorded with a SIS Morada digital camera using iTEM software (Olympus Soft Imaging Solutions GmbH, Munster, Germany). Bacteriophage morphology and related taxonomic assignment were confirmed following the guidelines set by the International Committee on Taxonomy of Viruses (http://www.ictvonline.org/; 39).

Phage stability

Phage stability in SM buffer was determined by measuring the EOP after incubation at different temperatures (21°C for 24 h, 21°C for 7 days, 37°C for 24 h, 4°C for >1 month, and 4°C for >1 month with chloroform) and at different pH levels (pH 3, 6, 7, and 8; 4 h at room temperature). Rate of adsorption and one-step growth curves (latent period and burst size) were calculated according to established protocols (37).

Results

CP-*K. pneumoniae* population

From a total of 21 isolates sequenced, 18 ESBL *K. pneumoniae* belonged to CG258 (S1 Table). This set contained two ST11 representatives, one ST512 (single locus variant of ST258 clade 2) (7), one ST1199 (single SNP variant of ST258 clade 1) (40), and 14 ST258 strains, segregated further into two distinct lineages, known as clade 1 (n=13) and clade 2

(n=2) and distinguished by approx. 40 non-recombinant core SNPs (Fig 1). Strains belonging to clade 1 dominated this population (86%), reflecting the epidemiology of a recent Australian outbreak (9). Inspection of the most common *K. pneumoniae* virulence-associated mobile genetic element, ICE*Kp*, revealed that all strains belonging to ST11 carried an ICE*Kp3* associated with a yersiniabactin operon *ybt* 9, while all ST258 and the ST1199 strain carried an ICE*Kp2* associated with a yersiniabactin operon *ybt* 13, a combination usually only found in clade 1 isolates (Fig. 1) (clade 2 isolates most commonly being associated with ICE*Kp10* with *ybt* 17) (41).

CP-*K. pneumoniae* CG258 bacteriophage susceptibility

Screening for bacteriophage susceptibility identified at least one phage with strong activity against each strain of CP-*K. pneumoniae* CG258, with potential therapeutic value (Fig. 2). A correlation pattern between specific regions of differences (RODs) and bacteriophage susceptibility was evident in our dataset (Fig.2). Phage host range profiles largely grouped according to capsular types, KL15-1 (ST11), KL106-D1 (ST258 clade 1) and KL107 (ST258 clade 2 and ST512) each presenting unique patterns of phage susceptibility with few examples of cross-reactivity. Phage AmPh_EK29 was able to lyse both a subset of clade 1 isolates (KL106-D1 and D2) and JIE2793 (ST512; KL107), while phages JIPh_Kp122 and JIPh_Kp127 had preferential activity against few specific clade 1 isolates (Fig. 2).

Within the clade 1 set, this capsule-specificity of the phage tested was further confirmed by the resistance to lysis by the capsular variant isolates JIE2783 and JIE4282 (see below) which were resistant to the majority of the bacteriophages capable of lysing other clade 1 strains. *K. pneumoniae* phages AmPh_EK52 and JIPh_130 were effective in clearing most CP-*K. pneumoniae* ST258 clade 1 specifically, except for the ones carrying capsular variants. There

296 was no cross-activity of any of these clade-specific phages against any of the control species
297 tested (*E. coli, E. faecium, P. aeruginosa, S. aureus* or *S. pseudintermedius*).
298
299 **ESBL *K. pneumoniae* CG258 variable accessory genome**
300 A comprehensive analysis of the pan-genome using Roary showed that the CG258 accessory
301 genome was composed of 1,682 accessory genes, which could be further classified into
302 discrete RODs (regions of difference) based on their function and contiguity as follows: 33
303 capsule-associated genes, 334 phage-related genes, 622 plasmid-related genes, 328 RODs-
304 associated genes, and 365 other genes that could not be assigned unambiguously to the
305 aforementioned categories (ROD were numbered and are summarised in S2 Table). The
306 presence-absence profile of these RODs appeared to reflect the ST and/or clade of the
307 isolates, in particular for prophage- and plasmid-related regions (Fig. 2ABD; S2 Table).
308 Notwithstanding this, there was a substantial degree of intra-clade variation within the ST258
309 clade 1 isolates in the RODs corresponding to the *cps* polysaccharide capsule encoding locus,
310 plasmid content and other RODs (Fig. 2ABD; S2 Table).
311
312 *Polysaccharide capsule synthesis locus*
313 Bioinformatic analysis of the genomic region between the *galF* and *ugd* genes revealed
314 distinct capsular types within the same ST, with KL15 and KL103 found in ST11 strains, and
315 KL106 and KL107 found in ST258 clade 1 and 2, respectively. The *cps* locus is a well-
316 known recombination hot-spot in *K. pneumoniae* and further inspection of our clade 1
317 isolates revealed the presence of three variants of the previously described KL106-D1
318 arrangement (40,42) (Fig. 2; Fig. 3a). Variants were due to insertion of IS*Kpn26* (IS5 family)
319 in two different locations within the *cps* locus: (i) within the *wcaJ* gene (KL106-D2 in
320 JIE2783 and KL106-D4 in JIE4282) and (ii) within the acyltransferase encoding gene

(KL106-D3 found in JIE4005, JIE4019, and JIE4020) (Fig. 3a). All insertions caused the interruption of open reading frames producing characteristic 4 bp duplications (Fig. 3a). Sequences representative of the three new *cps* locus variants were deposited in GenBank. Capsule production was significantly different among the CP-*K. pneumoniae* strains (ANOVA, $p < 0.001$) (Fig. 3b). Production was significantly decreased in JIE2783 and JIE4282 (capsular variants KL106-D2 and KL106-D4 respectively), while it appeared significantly higher in JIE3095 (Fisher's protected LSD, $p < 0.05$), though in this latter strain no sequence variation in the capsule encoding locus was identified.

In our set of strains, an evident association was observed between bacteriophage susceptibility profiles and capsular locus variants (Fig. 2) with very few of the tested phage (4/65; ~ 6%) showing any cross-clade specificity. This correlation between capsule type and phage susceptibility held when a clade 1-specific (KL106) (AmPh_EK52), a clade 2-specific (KL107) (AmPh_EK38) and a phage with inter-clade range (JIPh_Kp122) were blind-tested on a larger panel (n=48) of CP-*K. pneumoniae* isolates from Europe (S3 Table).

Other cell surface structures

In contrast to the variable *cps* locus structure, gene content and arrangement in the LPS encoding loci was well conserved in our isolates. Accordingly, lipopolysaccharide profiles from silver staining showed a high degree of homogeneity with no significant differences in the O antigens of the short, long or intermediate chains (S1 Fig.). The lipid A component of JIE4282 differed in size from all other *K. pneumoniae* CG258 (S1 Fig.). According to Kleborate (26), JIE4282 is missing the *wbbM* gene encoding a glycosyltransferase required for *d*-galactan I biosynthesis (43). Of note, although JIE2793 (ST512) is missing one of the hypothetical proteins (*glmA*) in the O antigen operon (O2v2 type), this had no observable impact on the observed LPS profile.

Biofilm production levels also differed (unbalanced ANOVA, $p < 0.001$) and were found to be significantly higher in JIE2793, JIE3095 and JIE4282 compared to all others (Fisher's protected LSD, $p < 0.05$) (S2 Fig.). Variable levels could be attributable to a number of factors, including variations associated with fimbrial genes. For instance, the *ecp* fimbrial operon associated with ROD-6 is missing in JIE4005, JIE4019, and JIE4020 (all KL106-D3) and in JIE4046 (S2 Table). JIE3095 also harbours a large 124 kb recombinant region which encompasses some fimbrial proteins. Limited capsule production (*e.g.* JIE4282) and expression of O-antigen variants (*e.g.* JIE2793) could also affect adhesion necessary for biofilm establishment. These variations in cell surface structures could be implicated with bacteriophage susceptibility, but no immediate correlation with phage host range was actually observed.

Prophages

The prophage content of clade 1 isolates was also variable. For example, a common prophage found in all other clade 1 representatives (designated phage 1 in S2 Table, with homology to viruses of the *Myoviridae* family) was truncated in JIE3095, missing most of the tail assembly module (Fig. 2). In JIE4005, 4019, and 4020 an additional unique prophage sequence was identified (Fig. 2; S2 Table). Based on prophage profile, clade 1 isolates could be subdivided into groups with broad association with bacteriophage susceptibility (Fig. 2).

Plasmids

All *K. pneumoniae* were MDR, and all ST258 isolates with the exception of JIE2740 carried the bla_{KPC} gene (Table 2; S4 Table). The overall plasmid signature for each ST258 clade was unique and remarkably uniform (Table 2; S3 Fig.). As expected (8,16), the bla_{KPC} allele 2 (bla_{KPC-2}) was exclusively associated with ST258 clade 1, whilst bla_{KPC-3} was associated with

clade 2 isolates and the closely related ST512 strain (Table 2). These genes were found exclusively on large (>20 kb) plasmids and co-localized by Southern hybridization with the IncFIIK *rep* gene (Table 2; S3 Fig.). All isolates carried multiple genes conferring extended-spectrum β-lactam resistance (ESBL) other than bla_{KPC} (S4 Table). Antibiotic resistance genotypes determined by WGS analysis accounted for all resistance phenotypes determined by standard clinical screening (S4 Table).

Table 2. Plasmid content in sequenced *K. pneumoniae* CG258 isolates.*

Isolate^	ST258 clade	bla_{KPC} allele	Plasmid replicons†	Plasmid sizes (kb)**	
				20-100	>100
2487	na (ST11)	none	IncFIIK IncF-like ColE-like	nd	nd
2713	na (ST11)	none	IncFIIK ColE-like	nd	nd
2709	1	2	IncFIIK IncFIB-pKpQil-like IncX3 ColE-like	none	242.5; **104.5**
2733	1	2	IncFIIK IncX3 ColE-like	**43**	194
2740	1	none	IncFIIK IncX3 ColE-like	41	194
2771	1	2	IncFIIK IncFIB-pKpQil-like IncX3 ColE-like	41	**104.5**
2783	1	2	IncFIIK IncFIB-pKpQil-like IncX3 ColE-like	**43**	194; **104.5**
3095	1	2	IncFIIK IncX3 ColE-like	41	**165**
4005	1	2	IncFIIK IncX3 ColE-like	43	**160.5**
4019	1	2	IncFIIK IncF IncX3 ColE-like	43	**160.5**; 150
4020	1	2	IncFIIK IncX3 ColE-like	43	**160.5**
4046	1	2	IncFIIK IncX3 ColE-like	41	**160.5**
4203	1	2	IncFIIK IncX3 ColE-like	41	194; **104.5**
4282	1	2	IncFIIK IncFIB-pKpQil-like IncX3 ColE-like	41	200; **110**
4455	1	2	IncFIIK IncFIB-pKpQil-like IncX3 ColE-like	43	**160.5**
2793	na (ST512)	3	IncFIIK IncN ColE-like	58	**208**

continues next page

Table 2 continued

Isolate[^]	**ST258 clade**	***bla*$_{KPC}$ allele**	**Plasmid replicons**[†]	**Plasmid sizes (kb)**[**]	
				20-100	**>100**
4626	2	3	IncFIIK IncR IncX3 ColE-like	**43**	208; 121
4660	2	3	IncFIIK IncR IncX3 ColE-like	**43**	208; 121

[*]Data obtained from WGS analysis except where otherwise specified. [^], in bold isolates from same patient; [†], based on PlasmidFinder scheme implemented in BAP (30); [**], approximate sizes determined by S1-PFGE; nd, not determined. In bold, plasmids co-localizing with the *bla*$_{KPC}$ gene as identified by Southern blot hybridization. Underline, IncFIIK replicons detected in Southern blot hybridization.

Bacteriophage characterization

Among the tested phages, we identified five unique double-stranded DNA bacteriophages AmPh_EK29, AmPh_EK52, AmPh_EK80, JIPh_Kp122, and JIPh_Kp127 that selectively target CP-*K. pneumoniae* ST258 clade 1 isolates, some of which may have therapeutic potential (Table 3; Table 4). WGS of purified viral DNA produced 11,340 to 5,049,732 reads that *de novo* assembled into one contig in all instances (S1 Table). Bacteriophage genomes size varied between 40.7 and 169.3 kb and GC content was lower than 50% (i.e. host genome) in all except for AmPh_EK52 (GC% 52.9) (Table 4; Fig. 4). No lysogeny or virulence associated genes were identified in these bacteriophage sequences, consistent with their lytic nature and thus indicating their suitability for therapeutic use (Fig. 4). The high degree of sequence similarity (>95%) to characterized *K. pneumoniae*-specific phage in the NCBI database and TEM imaging indicated that each selected phage belonged to the order Caudovirales (Table 4; Fig. 4).

AmPh_EK29 and JIPh_Kp122 (*Myoviridae*-like), presented a prolate head (approx. 80 by 100 nm) and contractile tale (100 nm) (Table 4; Fig. 4a and 4d). AmPh_EK52 resembled *Podoviridae* phages for having a small thick non-contractile tail (approx. 20 nm long) and close homology to other members of this family including genome size of about 40 kb and

absence of t-RNAs in its genome (Table 4; Fig. 4b). JIPh_Kp127 and AmPh_EK80 were both T5-like *Siphoviridae* viruses with long thin non-contractile tails (Table 4; Fig. 4c and 4e). Screening of entries in the NCBI database by BLASTn identified close relatives of these bacteriophages but no identical sequences (Table 4), and in genome comparisons with best matching GenBank entries the modular structure and order were preserved in all cases, with the main regions of difference found in tail or tail-associated open reading frames (Fig. 4).

All five bacteriophages efficiently lysed target bacteria *in vitro* at high titre and in combination captured the entire ST258 clade 1 subset (Table 3). However, host ranges were unique for each phage. AmPh_EK80, JIPh_Kp122 and JIPh_Kp127 showed limited activity toward ST258 strains, bar a few specific targets that lysed at high titre, and often were associated with confluent lysis zones, indicating the possibility of lysis from without or abortive lysis (Table 3). All phages were highly stable in SM buffer maintaining high titre at a range of temperatures (4°C for >1 month; 21°C for one week; 37°C for 24 h) and pH levels (pH 3, 6, 7 and 8 for 4 h). Exposure to chloroform at 4°C decreased the stability of AmPh_EK29 by 2-3 orders of magnitude, but had no effect on the stability of the remaining four phages. One-step growth curves revealed latent periods of 80-250 min and burst sizes between 12 and 500 pfu/cell (Table 4; S4 Fig.). Growth curves for AmPh_EK80 and JIPh_Kp127 were comparable. Phage JIPh_Kp122 had the shortest latent time (80 min), while AmPh_EK52 had the shortest burst time (30 min).

Accession number(s) The Illumina sequencing datasets of all *K. pneumoniae* isolates obtained in this work were deposited in SRA (NCBI) database under Bioproject PRJNA529495. The *cps* locus variants were deposited in the GenBank database (NCBI) under accessions: XXXXX for KL106-D2 (JIE2783), XXXXX for KL106-D3 (JIE4019), and XXXXX for KL106-D4 (JIE4282). The complete annotated genomes of *K. pneumoniae*

phages AmPh_EK29, AmPh_EK52, AmPh_EK80, JIPh_Kp122, and JIPh_Kp127 were also deposited in the GenBank database (NCBI) under accession numbers XXXXX, XXXXX, XXXXX, XXXXX and XXXXX, respectively.

Discussion

The increasing challenges posed by the rise of antibiotic resistance in human pathogens have revitalized interest in the use of bacteriophage for the treatment of bacterial infections (10-12). Among MDR pathogens, CP-*K. pneumoniae* is a serious clinical concern, as both gut colonizer and agent of severe sepsis when invading sterile body sites (1,44). In this study, ST258 isolates were predominant in the local clinical CP-*K. pneumoniae* population, with overrepresentation of ST258 clade 1 (KL106-D1), reflecting the epidemiology of a recent local outbreak (9). The incidence of bla_{KPC} in Australia has been rather limited when compared to its dissemination in other countries (16). The Western Sydney Health district is one of the largest in the country and we could only identify 0.2% CP-*K. pneumoniae* with bla_{KPC} over the past six years in its diagnostic enterobacterial collection (JIE), which is in line with reported national frequencies (45). However, as our own data confirms, tracking of these pathogens remains paramount due to the consistent association of multidrug resistance with mobilizable elements (16,36).

In the *K. pneumoniae* genome, the *cps* locus, encoding the capsular polysaccharide outer layer, is a recognized recombination hotspot, responsible for the diversification of clonal lineages, particularly within the CG258 group (6-8), and over 100 capsular types have been identified in this species (42,46). The capsule is a complex structure of repeating sugar subunits that protects the cell from external threats (including phage attack) and enhances *K. pneumoniae* virulence, being implicated in resistance to host defence mechanisms, immune evasion, adherence, and biofilm formation (47). *K. pneumoniae* types produce capsule in

varying degrees and hypermucoviscosity (excessive capsule production) has been linked to increased virulence (47). In our study, we identified three novel capsular variants in ST258 clade 1 due to ISKpn*26* insertion in the *cps* locus, and responsible for the intra-clade variation in our population which correlated remarkably well with reduced host range for many of the clade 1 infecting phages. ISKpn*26* has been recently implicated in unique recombination events in MDR *K. pneumoniae* (48) and may be worthwhile tracking. Two of the variants (D2 and D4), where the IS interrupted the *wcaJ* gene, were found in isolates with reduced capsule production in accordance with studies demonstrating that disruption of this gene has a detrimental effect on capsule content (46,49). The positive correlation between phage resistance and reduced virulence has been previously observed in the adaptive interplay between *Klebsiella* and its bacteriophages, with the isolation of less virulent bacterial mutants, and may be an expected trade-off in their evolutionary arms race (50). It is therefore not surprising that co-evolution of specific phage with this host resulted in the narrow host ranges observed in this study, a characteristic that could be exploited for therapeutic purposes, as specific recognition by phage receptors (tail spikes) of complementary anti-receptors on the host cell surface is at the core of phage lytic activity (51,52).

Comparative analysis of genomic data and bacteriophage susceptibility profiles showed a marked association between the host range of most of the tested viruses with capsular type. Very few examples of cross-reactivity were observed among *K. pneumoniae* CG258 phages that lysed different capsular types with high specificity. Host range was further restricted within clades, i.e. no phage lysing ST258 clade 1 effectively also infected clade 2 isolates, and *vice versa*. Phage AmPh_EK29 as a notable exception lysed only a subset of ST258 clade 1 strains (capsular type KL106), but also the ST512 representative isolate (capsular type KL107), indicative of a unique receptor specificity for this virus. JIPh_Kp122 and JIPh_Kp127 also showed broader spectra beyond capsular type, but lysed ST258 isolates

overall with poor efficiency, with evidence of abortive or passive lysis (53), perhaps limiting their therapeutic value. Capsule-targeting bacteriophages have been shown to be effective against *K. pneumoniae* of different capsular type *in vitro* and *in vivo*, along with depolymerases, which are enzymes produced by phages that cleave glycosidic bonds disrupting capsule integrity (54,55). Depolymerase activity was originally shown to be random, however recent work demonstrated depolymerase specificity toward certain K types (56,57), an indication that enzymatic degradation may be causing the patterns of lytic activity observed in our study. We did not find immediate evidence that all our phage encoded these proteins and, for the ones for which a depolymerase-related phenotype was observed, we could not readily identify the putative coding regions, except for AmPh_EK52.

Even though the attachment mechanisms or receptor binding sites for any of the tested phages were not investigated, host range matching with genomic data allowed for the selection of a number of viruses with unique characteristics that could be further examined for therapeutic applications. All the phages sequenced in this work were highly homologous to previously characterized lytic viruses shown to be effective against specific *K. pneumoniae* STs. However, none matched the specificities (host ranges) against ST258 of our viruses. Comparative analysis of phage genomes identified preferential loci of variability, mostly related with tail fibres or tail-associated genes, likely the primary receptors responsible for each specific host range. The different but complementary host ranges of the phages characterized here may indicate differences in receptor specificity (see AmPh_EK52 versus JIPh_Kp122 for example) and could lead to successful therapeutic combinations (high activity, poor resistance development).

The complex dynamic interactions of phages and their hosts and the co-evolution mechanisms at play have undermined the direct prediction of phage susceptibility from host genomics, even when dealing with clonal populations. Though capsular variation was at the

core of the divergence of isolates within this subset, fine genomic diversity in our ST258 clade 1 strains was also associated with other elements such as prophage content, porin defects, and plasmid content. The ST258 clade 1 subset in this study roughly divided into three subgroups based on prophage profile, further evidence of the genome plasticity in these species (58). Prophage content was likely implicated in resistance to some of the tested phages (e.g. AmPh_EK29) and must also be considered when designing optimal therapeutic mixes. Variability in the lipid A core of the LPS surface layer could also contribute phage resistance in JIE4282, as this is generally highly conserved within *K. pneumoniae*, making it an important determinant for host receptor recognition (51,52). Other differences (lack of fimbrial locus in JIE4046, different plasmid content (JIE4005, 4019 and 4020), *ompK36* variants etc.) seemed not to directly impact the susceptibility patterns to most of the tested phages. However, these all have the potential to affect viral host range and synergy when phages are to be used for therapeutic cocktails preparation (51,52).

The unique relationship between *K. pneumoniae* and its phages highlighted in this study is very important for future progress in therapeutic applications specifically targeting ST258 isolates or other hypervirulent types Better bioinformatic tools and larger well characterized microbial collections may allow for better definition of predictive algorithms. As we have demonstrated here capsular variation is critical to phage susceptibility and must be considered when designing effective therapeutics against this pathogen.

REFERENCES

[1] Podschun, R. and U. Ullmann (1998). *Klebsiella* spp. as nosocomial pathogens: epidemiology, taxonomy, typing methods, and pathogenicity factors. Clinical microbiology reviews 11(4): 589-603.

[2] World Health Organization. Antimicrobial Resistance Global Report on Surveillance: 2014 Summary; WHO: Geneva, Switzerland, 2014.

[3] Centers for Disease C. Prevention Vital signs: carbapenem-resistant *Enterobacteriaceae*. (2013). MMWR Morb Mortal Wkly Rep 62:165–70.

[4] Nordmann, P., et al. (2009). The real threat of *Klebsiella pneumoniae* carbapenemase-producing bacteria. The Lancet Infectious Diseases 9(4): 228-236.

[5] Pitout, J. D. D., et al. (2015). Carbapenemase-producing *Klebsiella pneumoniae*, a key pathogen set for global nosocomial dominance. Antimicrobial agents and chemotherapy 59(10): 5873-5884.

[6] Deleo, F. R., et al. (2014). Molecular dissection of the evolution of carbapenem-resistant multilocus sequence type 258 *Klebsiella pneumoniae*. Proceedings of the National Academy of Sciences 111(13): 4988-4993.

[7] Holt, K. E., et al. (2015). Genomic analysis of diversity, population structure, virulence, and antimicrobial resistance in *Klebsiella pneumoniae*, an urgent threat to public health. Proceedings of the National Academy of Sciences of the United States of America 112(27): E3574-E3581.

[8] Wyres, K. L., et al. (2015). Extensive Capsule Locus Variation and Large-Scale Genomic Recombination within the *Klebsiella pneumoniae* Clonal Group 258. Genome biology and evolution 7(5): 1267-1279.

539 [9] Kwong, J. C., et al. (2018). Translating genomics into practice for real-time surveillance
540 and response to carbapenemase-producing *Enterobacteriaceae*: evidence from a complex
541 multi-institutional KPC outbreak." PeerJ 6: e4210.
542 [10] Abedon, S. T., et al. (2011). Phage treatment of human infections. Bacteriophage 1(2):
543 66-85.
544 [11] Chan, B. K., et al. (2013). Phage cocktails and the future of phage therapy. Future
545 Microbiology 8(6): 769-783.
546 [12] Ly-Chatain, M. H. (2014). The factors affecting effectiveness of treatment in phages
547 therapy. Frontiers in Microbiology 5: 51-51.
548 [13] Kęsik-Szeloch, A., et al. (2013). Characterising the biology of novel lytic bacteriophages
549 infecting multidrug resistant *Klebsiella pneumoniae*. Virology journal 10: 100-100.
550 [14] Thiry, D., et al. (2019). New Bacteriophages against Emerging Lineages ST23 and
551 ST258 of *Klebsiella pneumoniae* and Efficacy Assessment in Galleria mellonella Larvae.
552 Viruses 11(5): 411.
553 [15] Ciacci, N., et al. (2018). Characterization of vB_Kpn_F48, a newly discovered lytic
554 bacteriophage for *Klebsiella pneumoniae* of sequence type 101. Viruses 10(9): 482.
555 [16] Partridge, S. R., et al. (2015). Emergence of blaKPC carbapenemase genes in Australia.
556 International Journal of Antimicrobial Agents 45(2): 130-136.
557 [17] Shoma, S., et al. (2014). Characterization of multidrug-resistant *Klebsiella pneumoniae*
558 from Australia carrying blaNDM-1. Diagnostic Microbiology and Infectious Disease 78(1):
559 93-97.
560 [18] O'Tootle G. A. and R. Kolter (1998). Initiation of biofilm formation in *Pseudomonas*
561 *fluorescens* WCS365 proceeds via multiple, convergent signalling pathways: a genetic
562 analysis. Molecular Microbiology 23(8): 449-61.

563 [19] Domenico, P., et al. (1989). Reduction of capsular polysaccharide production in
564 *Klebsiella pneumoniae* by sodium salicylate. Infection and Immunity 57(12): 3778-3782.
565 [20] Hsu, C.-R., et al. (2016). Identification of a capsular variant and characterization of
566 capsular acetylation in *Klebsiella pneumoniae* PLA-associated type K57. Scientific Reports
567 6: 31946.
568 [21] Hitchcock, PJ, Brown, TM. (1983). Morphological heterogeneity among Salmonella
569 lipopolysaccharide chemotypes in silver-stained polyacrylamide gels. Journal of
570 Bacteriology, 154(1), pp. 269-277
571 [22] Chevallet, M, Luche, S, Rabilloud, T. (2006). Silver staining of proteins in
572 polyacrylamide gels. Nature Protocols, 1(4), pp. 1852-1858
573 [23] Fajardo-Lubián, A., et al. (2019). Host adaptation and convergent evolution increases
574 antibiotic resistance without loss of virulence in a major human pathogen. PLOS Pathogens
575 15(3): e1007218.
576 [24] Bankevich, A., et al. (2012). SPAdes: a new genome assembly algorithm and its
577 applications to single-cell sequencing. Journal of computational biology: a journal of
578 computational molecular cell biology 19(5): 455-477.
579 [25] Seemann, T., et al. Nullarbor Github https://github.com/tseemann/nullarbor
580 [26] Lam, M., et al. (2018) Kleborate: comprehensive genotyping of *Klebsiella pneumoniae*
581 genome assemblies. https://github.com/katholt/Kleborate.
582 [27] Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-
583 analysis of large phylogenies. Bioinformatics 30(9): 1312-1313.
584 [28] Croucher, N. J., et al. (2015). Rapid phylogenetic analysis of large samples of
585 recombinant bacterial whole genome sequences using Gubbins. Nucleic acids research 43(3):
586 e15.

587 [29] Page, A. J., et al. (2015). Roary: rapid large-scale prokaryote pan genome analysis.
588 Bioinformatics 31(22): 3691-3693.
589 [30] Thomsen, M. C. F., et al. (2016). A bacterial analysis platform: an integrated system for
590 analysing bacterial whole genome sequencing data for clinical diagnostics and surveillance.
591 PLOS ONE 11(6): e0157718.
592 [31] Arndt, D., et al. (2016). PHASTER: a better, faster version of the PHAST phage search
593 tool. Nucleic acids research 44(W1): W16-W21.
594 [32] Sullivan, M. J., et al. (2011). Easyfig: a genome comparison visualizer. Bioinformatics
595 (Oxford, England) 27(7): 1009-1010.
596 [33] Agyekum, A., et al. (2016). *bla*CTX-M-15 carried by IncF-type plasmids is the
597 dominant ESBL gene in *Escherichia coli* and *Klebsiella pneumoniae* at a hospital in Ghana.
598 Diagnostic Microbiology and Infectious Disease 84 (4):328-333.
599 [34] Barton, B. M., et al. (1995). A general method for detecting and sizing large plasmids.
600 Analytical Biochemistry 226(2): 235-240.
601 [35] Bradford P.A., et al. (2004). Emergence of carbapenem-resistant *Klebsiella* species
602 possessing the class A carbapenem-hydrolyzing KPC-2 and inhibitor-resistant TEM-30 β-
603 lactamases in New York City. Clinical Infectious Diseases 39:55–60.
604 [36] Villa L, et al. (2010). Replicon sequence typing of IncF plasmids carrying virulence and
605 resistance determinants. Journal of Antimicrobial Chemotherapy 65:2518–29.
606 [37] Clokie MRJ, Kropinski A, Eds. Methods and Protocols, Volume 1: Isolation,
607 Characterization, and Interactions. Springer Protocols 2009.
608 [38] Brettin, T., et al. (2015). RASTtk: A modular and extensible implementation of the
609 RAST algorithm for building custom annotation pipelines and annotating batches of
610 genomes. Scientific Reports 5: 8365.

611 [39] Adriaenssens, E. M., et al. (2015). Integration of genomic and proteomic analyses in the
612 classification of the Siphoviridae family. Virology 477: 144-154.
613 [40] Wyres, K.L., et al. (2016). Identification of *Klebsiella* capsule synthesis loci from whole
614 genome data. Microbial Genomics 2:e000102.
615 [41] Lam, M. M. C., et al. (2018). Genetic diversity, mobilisation and spread of the
616 yersiniabactin-encoding mobile element ICEKp in *Klebsiella pneumoniae* populations.
617 Microbial genomics 4(9): e000196.
618 [42] Follador, R., et al. (2016). The diversity of *Klebsiella pneumoniae* surface
619 polysaccharides. Microbial genomics 2(8): e000073-e000073.
620 [43] Kos, V., and C. Whitfield (2010) A membrane-located glycosyltransferase complex
621 required for biosynthesis of the D-galactan I lipopolysaccharide O antigen in *Klebsiella*
622 *pneumoniae*. Journal of Biological Chemistry 285(25):19668-87.
623 [44] Bengoechea, J. A., and J. S. Pessoa (2018). *Klebsiella pneumoniae* infection biology:
624 living to counteract host defences. FEMS Microbiology Reviews 43(2):123-144.
625 [45] Bell, J.M., et al. (2018). Australian Group on Antimicrobial Resistance (AGAR)
626 Australian Gram-negative Sepsis Outcome Programme (GNSOP) Annual Report 2016.
627 Communicable Diseases Intelligence 42:S2209-6051(18)00017-9.
628 [46] Pan, Y.-J., et al. (2015). Genetic analysis of capsular polysaccharide synthesis gene
629 clusters in 79 capsular types of *Klebsiella* spp. Scientific Reports 5: 15573.
630 [47] Ko, K. S. (2016). The contribution of capsule polysaccharide genes to virulence of
631 *Klebsiella pneumoniae*. Virulence 8(5): 485-486.
632 [48] Pitt, M. E., et al. (2018). Multifactorial chromosomal variants regulate polymyxin
633 resistance in extensively drug-resistant *Klebsiella pneumoniae*. Microbial genomics 4(3):
634 e000158.

635 [49] Dorman, M. J., et al. (2018). The Capsule Regulatory Network of *Klebsiella pneumoniae*
636 Defined by density-TraDISort. mBio 9(6): e01863-01818.
637 [50] Gu, J., et al. (2012). A Method for Generation Phage Cocktail with Great Therapeutic
638 Potential. PLOS ONE 7(3): e31698.
639 [51] Bertozzi Silva, J., et al. (2016). Host receptors for bacteriophage adsorption. FEMS
640 Microbiology Letters 363(4).
641 [52] Nobrega, F. L., et al. (2018). Targeting mechanisms of tailed bacteriophages. Nature
642 Reviews Microbiology 16(12): 760-773.
643 [53] Abedon, S. T. (2011). Lysis from without. Bacteriophage 1(1): 46-49. (Abedon 2011)
644 [54] Hsu, C.-R., et al. (2013). Isolation of a bacteriophage specific for a new capsular type of
645 *Klebsiella pneumoniae* and characterization of its polysaccharide depolymerase. PLOS ONE
646 8(8): e70092.
647 [55] Latka, A., et al. (2017). Bacteriophage-encoded virion-associated enzymes to overcome
648 the carbohydrate barriers during the infection process. Applied Microbiology and
649 Biotechnology 101(8): 3103-3119.
650 [56] Pires, D.P., et al. (2016). Bacteriophage-encoded depolymerases: their diversity and
651 biotechnological applications. Applied Microbiology and Biotechnology 100: 2141.
652 [57] Pan, Y.-J., et al. (2017). *Klebsiella* phage ΦK64-1 encodes multiple depolymerases for
653 multiple host capsular types. Journal of Virology 91(6): e02457-02416.
654 [58] Canchaya, C., et al. (2004). The impact of prophages on bacterial chromosomes.
655 Molecular Microbiology 53(1): 9-18.

656

657 **Acknowledgements**

658 The authors would like to acknowledge Alicia Arnott, Nathan Bachmann, Chayanika Biswas,
659 Rajat Dhakal, Elena Martinez, Ranjeeta Menon, Rebecca Rockett, Rosemarie Sadsad,

660 Verlaine Timms, Qinning Wang and Vitali Sintchenko at the Pathogen Genomics Unit,
661 Centre for Infectious Disease and Microbiology –Public Health, Westmead Hospital, for their
662 assistance with whole genome sequencing; Dongwei Wang for her assistance with TEM;
663 Sydney Water, Murray McDermott and Graziano (Rowland Village), and Lee Thomas for
664 their help with specimen collection for bacteriophage isolation.

665

666 **Funding**

667 This project was funded by the National Health and Medical Research Council (NHMRC;
668 GRP1107322). The Westmead Scientific Platforms are supported by the Westmead Research
669 Hub, the Cancer Institute New South Wales, the NHMRC and the Ian Potter Foundation.

670

671 **Conflicts of Interest**

672 None to declare.

673

Figures and Tables

Table 3. Efficiency of plating of selected CP-K. pneumoniae ST258 clade 1 bacteriophages*.

Isolate (JIE)	ST	Clade	AmPh_EK29		AmPh_EK52		AmPh_EK80		JIPh_Kp122		JIPh_Kp127	
2713	ST11	NA	none	4	none	4	none	4	***	3	***	3
2487	ST11	1	none	4	none	4	none	4	***	3	***	3
4203	ST258	1	none	4	$1.0x10^9$ (125)	1	none	4	***	3	***	3
4046	ST258	1	$1.1x10^9$ (55)	1	$8.0x10^8$ (100)	1	***	3	***	3	***	3
4020	ST258	1	none	4	$4.0x10^9$ (500)	1	***	3	***	3	***	3
4019	ST258	1	none	4	$2.6x10^9$ (325)	1	***	3	***	3	***	3
4005	ST258	1	none	4	$1.6x10^9$ (200)	1	***	3	***	3	***	3
3095	ST258	1	$4.6x10^8$ (23)	1	$8.0x10^8$ (100)	1	***	3	***	3	***	3
2709	ST258	1	**$2.0x10^9$ (100)**	**1**	**$8.0x10^8$ (100)**	**1**	$4.0x10^8$ (15)	1	***	3	***	3
2733	ST258	1	$8.0x10^8$ (40)	1	$6.0x10^8$ (75)	1	$4.0x10^8$ (15)	1	***	3	***	3
2740	ST258	1	$4.0x10^8$ (20)	1	$2.6x10^9$ (325)	1	none	4	***	3	***	3
2771	ST258	1	$1.8x10^9$ (90)	1	$4.0x10^8$ (50)	1	***	3	***	3	***	3
2783	ST258	1	$2.4x10^9$ (120)	1	none	4	$6.0x10^9$ (100)	1	$2.5x10^9$ (313)	1	$1.5x10^7$ (188)	1
4455	ST258	1	$2.0x10^9$ (100)	1	$2.5x10^8$ (31)	1	***	3	***	3	***	3
4282	ST258	1	none	4	none	4	**$6.0x10^9$ (100)**	**1**	**$8.0x10^8$ (100)**	**1**	**$8.0x10^6$ (100)**	**1**
2793	ST512	NA	$6.0x10^8$ (30)	1	none	4	none	4	***	3	***	3
4660	ST258	2	none	4	none	4	none	4	***	3	***	3
4626	ST258	2	none	4	none	4	none	4	***	3	***	3

*, in brackets percentage EOP when compared to amplification strain. ***, indicates clearing in bacterial lawn without single plaque formation (confluent lysis or abortive lysis). Bold, original amplification host. Scores indicate: 1, high titre lysis; 3 and 4, poor or no lysis.

Table 4. Characteristics of selected *K. pneumoniae* ST258 clade 1-specific bacteriophages.

Name (origin)	Size bp	GC %	Best NCBI match* (accession; bp)	Identity %	Coverage %	Classification	t-RNA	Depolymerase activity	latent period	burst time	burst size
AmPh_EK29 (Australia)	169353	40.8	vB_Kpn_F48 (MG746602; 170764)	98	95	*Myoviridae*	yes (7)	none detected	130	50	500
AmPh_EK52 (Australia)	40713	52.9	KP32 isolate 192 (MH172261; 40635)	93	93	*Podoviridae; Autographivirinae*	no (0)	yes	110	30	250
AmPh_EK80 (Australia)	112215	44.8	Sugarland (MG459987; 111103)	97	91	*Siphoviridae*; T5virus	yes (22)	yes	240	90	75
JIPh_Kp122 (Australia)	166475	39.6	JD18 (KT239446; 166313)	98	97	*Myoviridae*; *Tevenvirinae*; Jd18virus	yes (16)	none detected	80	60	12
JIPh_Kp127 (Australia)	113671	45.2	vB_Kpn_IME260 (KX845404.2; 123490)	94	92	*Siphoviridae*; T5virus	yes (22)	none detected	250	80	142

**K. pneumoniae* specific bacteriophages.

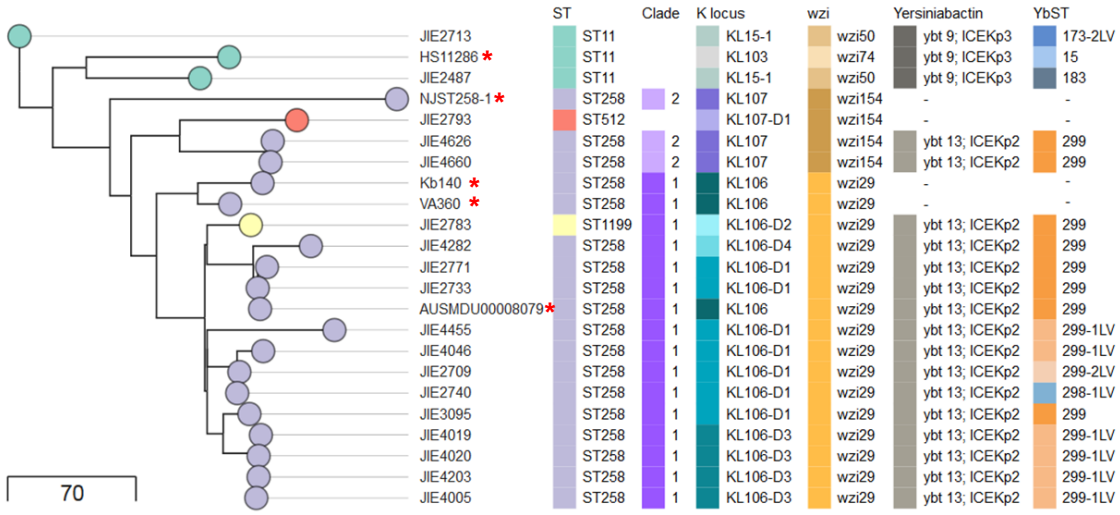

Figure 1. *K. pneumoniae* CG258 phylogeny. Recombination-free maximum-likelihood phylogenetic tree of 18 *K. pneumoniae* CG258 isolates and five publically available reference genomes (*). Metadata include MLST; CG258 clade; *cps* type (KL and *wzi*); Yersiniabactin, ICE*Kp* and ST (YbST). Tree scale corresponds to number of substitutions.

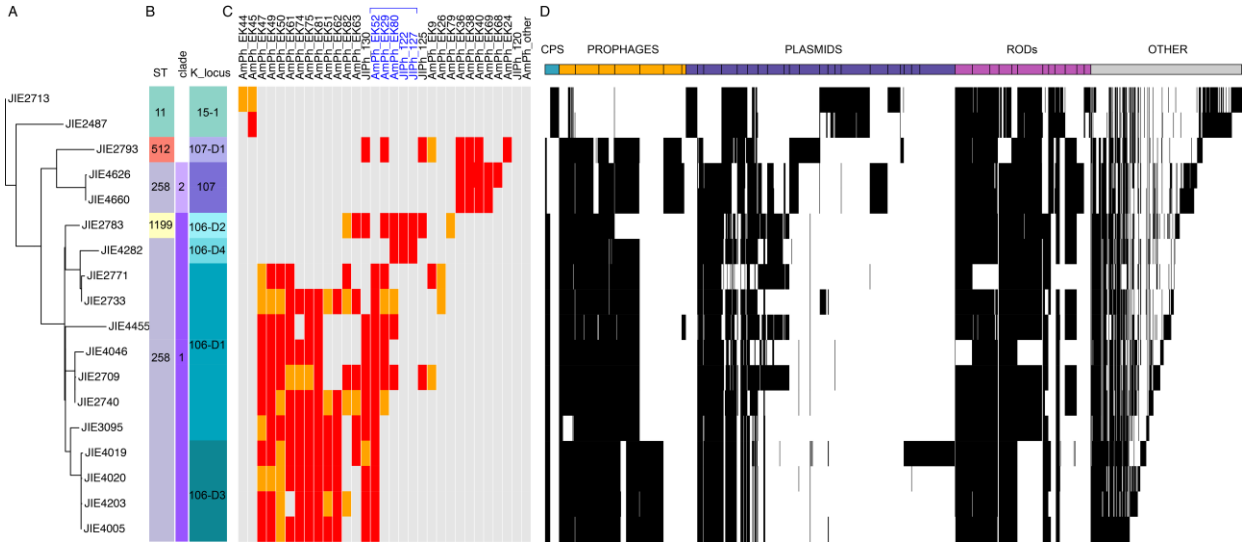

Figure 2. Phage susceptibility profiles and regions of differences. A) *K. pneumoniae* CG258 isolates phylogeny; **B)** metadata including ST, CG258 clade and K locus type; **C)** phage susceptibility profiles colour-coded based on therapeutic application potential as follows; grey, unsuitable for further testing (no lysis); orange, poor lytic activity; red, best candidates for further characterization (high lytic activity). Five phages (highlighted in blue) were selected for full characterization; **D)** regions of difference profiles identified using Roary (black, present). Genes corresponding to the accessory genome (n=1683) were reordered according to their synteny where possible, and classified according to functional categories into regions of difference as follows: CPS, capsule associated-genes; prophage-related regions; plasmid-related regions; other regions of differences (RODs) such as ICE elements; and other genes present in variable regions <10 consecutive genes. Details of regions are listed in Table S2.

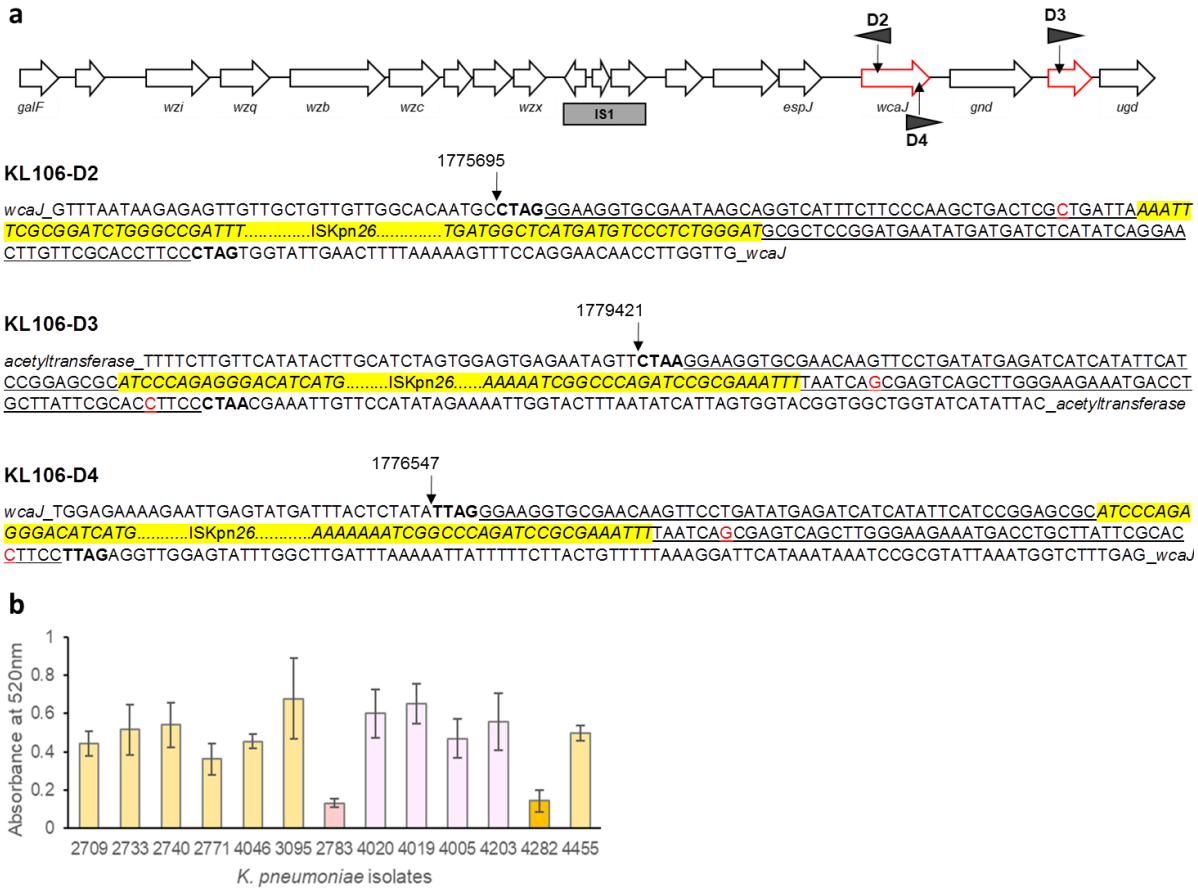

Figure 3. *cps* locus variants in *K. pneumoniae* ST258 clade 1 isolates. a) Schematic representing the *cps* locus gene arrangement in ST258 clade 1 isolates (KL106-D1; GenBank KR007677.1). Black triangles indicate ISKpn*26* insertions (directional) producing three variants (D2, D3 and D4). Drawing not to scale. Nucleotide sequence for ISKpn*26* insertions interrupting the *wcaJ* open reading frame (D2 and D4) and the acetyltransferase gene (D3) in the *cps* locus. In **bold**, insertion sequence (IS) direct repeats. Underlined, IS imperfect inverted repeats. 'Yellow highlight', indicates complete ISKnp*26* coding region. Nucleotide sequences of *cps* loci for variants KL106-D2 (JIE2783), KL106-D3 (JIE4005) and KL106-D4 (JIE4282) were deposited in GenBank. **b)** Capsule production assay. Capsule produced by ST258 clade 1 isolates was quantified based on uronic acid content (19,20). Results presented are average counts of three separate replicates.

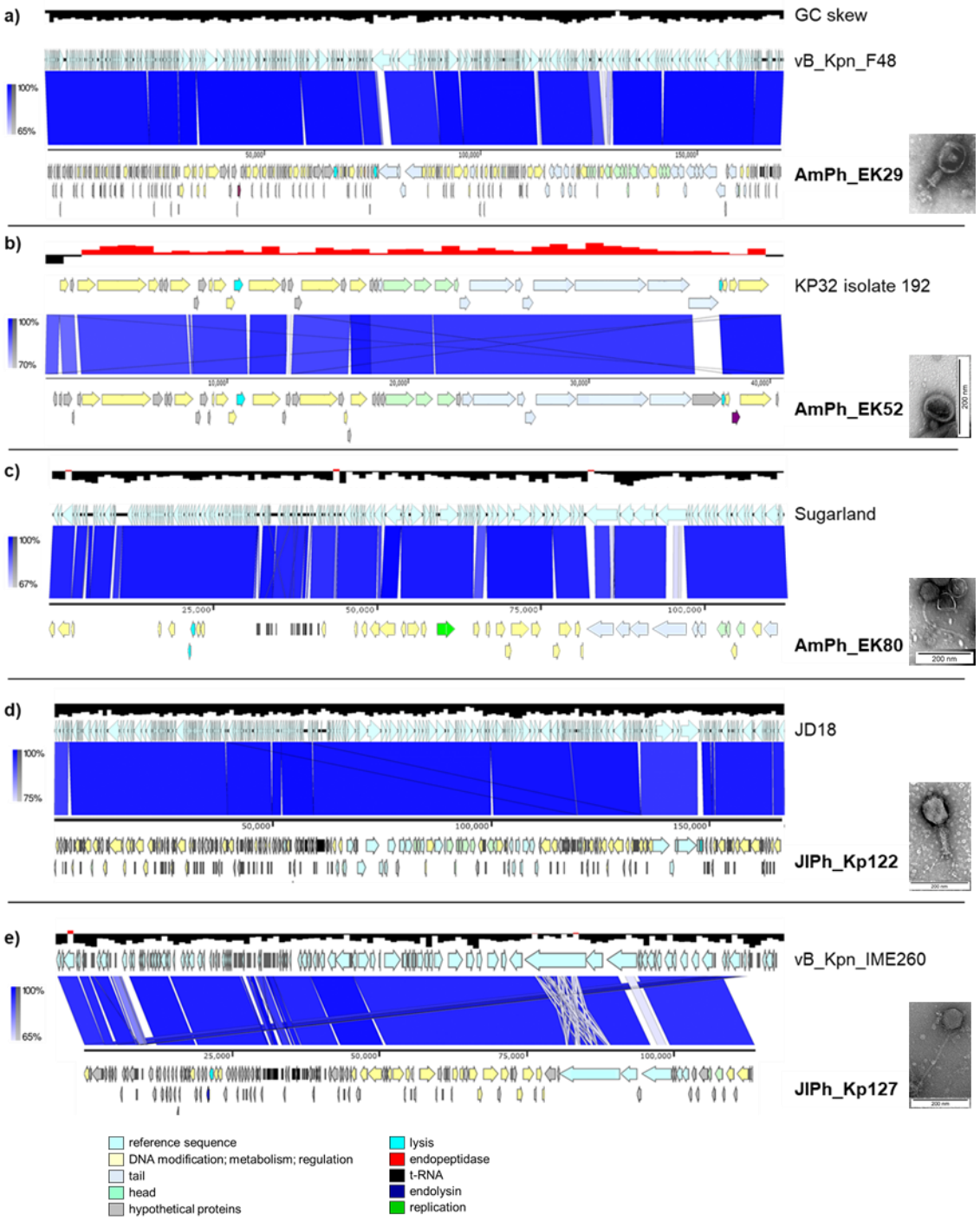

Figure 4. Comparative genome analysis of five sequenced *K. pneumoniae* ST258 bacteriophages. Schematics were produced using EasyFig (32) and show the structural organization of the selected phages compared to their best match (>95% nucleotide homology) reference in the GenBank database. Genes are color coded according to function. Phage morphology was captured using transmission electron microscopy.