Title: A One Health study of the genetic relatedness of *Klebsiella pneumoniae* and their mobile elements in the East of England

Authors: Catherine Ludden[1,2*#], Danesh Moradigaravand[2#], Dorota Jamrozy[3], Theodore Gouliouris[3,4,5], Beth Blane[3], Plamena Naydenova[3], Juan Hernandez-Garcia[3], Paul Wood[6], Nazreen Hadjirin[3], Milorad Radakovic[3], Charles Crawley[4], Nicholas M. Brown[4,5], Mark Holmes[3], Julian Parkhill[2], Sharon J. Peacock[1,2,3,4]

Affiliations:

[1] London School of Hygiene & Tropical Medicine, UK

[2] Wellcome Sanger Institute, UK

[3] University of Cambridge, UK

[4] Cambridge University Hospitals NHS Foundation Trust, UK

[5] Public Health England, UK

[6] University of Edinburgh, UK

[#]Equal contributions

Corresponding author: Catherine Ludden (catherine.ludden@lshtm.ac.uk)

Summary A One Health genomic survey of *K. pneumoniae* demonstrated transmission between patients and their ward environmental transmission, but no evidence to implicate livestock as a source of *K. pneumoniae* or their mobile genetic elements encoding antibiotic resistance.

ABSTRACT

Klebsiella pneumoniae is a human, animal and environmental commensal and a leading cause of nosocomial infections, which are often caused by multi-resistant strains that are challenging to treat. We conducted a One Health evaluation of putative sources of *K. pneumoniae* that are carried by, and infect hospital patients. This combined data from a six-month study on two haematology wards at Addenbrooke's Hospital, Cambridge, in 2015 to isolate *K. pneumoniae* from stool, blood and the environment, and a cross-sectional survey of *K. pneumoniae* from 29 livestock farms, 97 meat products, the hospital sewer and 20 municipal wastewater treatment plants in the East of England between 2014 and 2015. *K. pneumoniae* was isolated from stools of 17/149 (11%) patients and 18/922 swabs of their environment, together with one patient bloodstream infection during the study and 4 others over a 24-month period. Each patient carried one or more lineages that was unique to them, but two broad environmental contamination events and patient-environmental transmission were identified. *K. pneumoniae* was isolated from cattle and poultry, hospital sewage and 12/20 wastewater plants. There was low genetic relatedness between isolates from patients/their hospital environment versus isolates from elsewhere. Identical genes encoding cephalosporin resistance were carried by isolates from different reservoirs, but were carried on different plasmids by isolates from patients/their environment versus elsewhere. We identified no patient-to-patient transmission and no evidence for livestock as a source of *K. pneumoniae* infecting humans, but our findings reaffirm the importance of the hospital environment as a source of *K. pneumoniae* associated with serious human infection.

Keywords *Klebsiella pneumoniae*, One Health, genomics, antimicrobial resistance, surveillance

INTRODUCTION

Klebsiella pneumoniae is a major cause of nosocomial infections worldwide, the public health importance of which has been amplified by the increasing prevalence of multidrug resistant *K. pneumoniae* carriage and infection [1-3]. Resistance can be attributed to the global dissemination of extended-spectrum β-lactamase (ESBL)-producing *K. pneumoniae* and the subsequent emergence of isolates with resistance to the carbapenem drugs and colistin. Control of human *K. pneumoniae* infection requires an understanding of the potential sources from which infecting organisms are acquired. This is complex since *K. pneumoniae* can persist in a broad range of reservoirs including the hospital environment [4-7], retail meat [8, 9], livestock [10, 11] and wastewater [12, 13].

Attributing sources of *K. pneumoniae* that cause infection requires the sampling of multiple reservoirs in the same time and place, and comparing these using whole genome analyses. Previous genomic studies have largely focused on transmission in high-risk settings such as intensive care units. These have identified *K. pneumoniae* carriage as a significant risk factor for infection [14, 15]. Accurate interpretation of transmission also requires an understanding of whether specific reservoirs and individual samples contain more than one *K. pneumoniae* lineage. A previous study identified limited within-host diversity in 20% of patients based on sequencing of up to three *K. pneumoniae* colonies from 40 patients [15].

Here, we explore the genetic relatedness of *K. pneumoniae* isolated from the same and different reservoirs within a defined geographic region through an investigation of isolates from a patient cohort, their hospital environment, livestock, municipal wastewater and hospital sewage in the same geographic region of England.

MATERIALS AND METHODS

Patient recruitment and sampling

A prospective longitudinal study was conducted in two adult hematology wards at the Cambridge University Hospital NHS Foundation Trust (CUH) between May and November 2015. Patients were enrolled after informed written consent, after which stool samples were requested on the day of admission, every week thereafter and at discharge and cultured for *K. pneumoniae*. Environmental sampling for *K. pneumoniae* was performed throughout the study, details of which are provided in Supplementary Material. All blood cultures from patients on the two study wards between May 2014 and May 2016 were identified and stored isolates obtained. Wastewater was sampled from the main sewer of CUH on four spaced occasions between September 2014 and December 2015. A cross-sectional survey was conducted between June 2014 and January 2015 to isolate *K. pneumoniae* from raw and treated wastewater at 20 municipal wastewater treatment plants in the East of England. Ten plants were located downstream of acute hospitals and 10 did not directly receive hospital waste. A cross-sectional survey was conducted between August 2014 and April 2015 to isolate *K. pneumoniae* from livestock at 29 farms in the East of England (10 cattle (5 beef/5 dairy), 10 pig, and 9 poultry (4 chicken/5 turkey). Supplementary Material provide further details of the study design, sample collection and laboratory methods for bacterial culture, identification and antimicrobial susceptibility testing.

Whole-genome sequence analysis

Bacterial DNA was extracted using the QIAxtractor (QIAgen) and sequenced on an Illumina HiSeq2000 (Illumina, San Diego, CA, USA). *de novo* assembly of short read data was performed as previously described [16, 17], assemblies were annotated using Prokka [18] and a core genome alignment was produced by Roary. Further analysis was performed for ST307, ST268, ST6, ST34,

ST3010 and ST2585. Isolates belonging to these six STs were each mapped against a study isolate with the best N50 for that ST. Mapping was performed using SMALT (http://www.sanger.ac.uk/resources/software/smalt/) and recombination was removed using Gubbins [19]. SNPs were extracted using an in-house tool (www.github.com/sanger-pathogens/snp-sites). Maximum Likelihood trees were created using RAxML with 100 bootstraps and a midpoint root. Phylogenetic trees and associated metadata were visualised using iTOL[20] and Figtree (www.tree.bio.ed.ac.uk/software/figtree/). Identification of multilocus sequence types (STs), antimicrobial resistance determinants, virulence factors and plasmids was performed. A detailed description of the rationale for selecting isolates for sequencing and genomic analyses are provided in Supplementary Material.

RESULTS

Isolation of K. pneumoniae from stool, blood and the environment of a hospital cohort

174 of 338 patients (51%) admitted to the two study wards between May and November 2015 were recruited and agreed to provide stools for culture (Figure 1A). The 174 cases had a median stay of 19 days (interquartile range (IQR) 9 to 29 days), and were admitted a median of once (IQR 1 to 2, total 307 admissions). Stool samples were actually obtained from 149/174 patients (376 stools, median 3 (IQR 2-5) per case), with 101 patients providing two or more samples. *K. pneumoniae* was isolated from 23 stools from 17/149 (11%) patients, 3 of whom (2%) carried ESBL-producing *K. pneumoniae*, a similar rate to that reported previously [21]. 922 environmental swabs were taken from patient areas, medical equipment, and the wider environment over the 6-month study period. *K. pneumoniae* was isolated from 18 (2%) swabs from 7 different locations (3 single rooms/bathrooms during an initial point prevalence survey, and 3 single rooms/bathrooms plus a 2-bedded bay thereafter) each positive on one occasion. Six positive locations were in ward A and one was in ward B. This low rate of contamination is consistent with previous studies reporting 0-5% positivity during routine

surveillance [22, 23]. Bloodstream infection caused by *K. pneumoniae* occurred in 5 patients on the two study wards during a more extended 2-year period (May 2014 to May 2016), including one case during the 6-month study. Seven *K. pneumoniae* cultures from the 5 patients had been stored and were available for sequencing (Supplementary Table 1).

Genomic analysis of healthcare-associated isolates

We sequenced 122 *K. pneumoniae* colonies (termed isolates) picked from primary culture plates from 23 stools/17 participants (median 5 isolates per stool, range 1-15). These were assigned to 21 different sequence types (STs) (Figure 1B and Supplementary Table 1). Most patients (15/17) carried a single ST, with the two remaining cases each carrying 3 different STs (Supplementary Table 1). No two patients carried the same ST, indicating an absence of patient-to-patient transmission. Pairwise analysis of single nucleotide polymorphisms (SNP) in the core genome of isolates from the same patient/same ST demonstrated a median (range) difference of 1 (0-30) SNPs for the 17 patients.

We sequenced 24 isolates from 18 environmental swabs. These belonged to a more restricted population of 4 STs (Figure 1B and Supplementary Table 1). Integration of genetic and epidemiological data suggested two broad environmental contamination events. A cluster of 11 *K. pneumoniae* ST23 isolates were cultured from two adjacent single rooms and one more distant single room in Ward A, all positive on a single date in April 2014. Pairwise core genome SNP analysis demonstrated a median (range) difference of 2 (range 0-5) SNPs after removing two outliers ($\geq$ 45 SNPs different). A second cluster of 8 *K. pneumoniae* ST307 isolates were cultured from two adjacent single rooms in Ward A and 1 bedside in a 2-bedded bay in Ward B, all positive on a single date in Sept 2015. The 8 isolates were identical at the core genome level. In addition, *K. pneumoniae* ST268 and ST3021 were each isolated once from different single rooms. The ST3021 isolate was cultured

from the same room and day as an ST23 isolate, suggesting a wider contamination event and/or inadequate cleaning.

Comparison of isolates from stool and the environment showed that two STs (268 & 307) were identified from both sources (Figure 2A). ST268 was isolated from a single stool from a patient (D034) and their environment in August and October 2015, respectively. Patient D304 had 17 admissions to the haematology day unit and 2 transfers to Ward A during the period between the positive stool and environment sample. The most closely related stool-environmental ST268 isolate pair had highly similar core genomes (1 SNP different), which is consistent with recent patient-environment transmission. ST307 was isolated from stool in August 2015 from a patient (C029) in a single room on Ward B and 18 days later from the environment of two rooms in Ward A and one room in Ward B (a different room to the index case). The most closely related stool and environmental ST307 isolate pair were 64 core genome SNPs different (see Figure 3A for phylogeny), which is not consistent with recent patient-environment transmission.

We sequenced 16 *K. pneumoniae* isolates from 7 blood cultures/5 patients (1 isolate from each of the archived collection for 6 cultures/4 patients before or after the 6-month study, and 10 primary plate colonies from the case that occurred during the study). These belonged to five different STs, with one ST per patient. Four of the five STs were not identified in any other patient during the 6-month study, including the ST isolated from the patient who was bacteremic in this period. The exception was ST307, isolated from a patient (B024) with a bloodstream infection on two time points in September and December 2014 (at least 5 months before the prospective study began). Blood culture isolates from September and December differed by 4 SNPs, indicating relapse or reinfection with the same strain. Comparison between these and the prospective study isolates showed that the earlier blood culture isolates were related to the ST307 patient C029 stool isolates

(13 SNPs, range 12-15) and less closely to the ST307 environmental isolates (54 SNPs, range 52-57) (Figure 3A). This is indicative of either a persistent environmental reservoir of ST307 *K. pneumoniae* or on-going transmission between unsampled patients. Sequencing and pairwise core genome analysis of 10 colonies from a single blood culture from one patient (B022, ST6) demonstrated a median (range) difference of 1 (0-2) SNPs. This patient did not provide a stool sample.

Isolation of K. pneumoniae from livestock, retail meat and sewage

The number of sampling points at each farm was maximized by taking pooled samples containing up to 50 aliquots of freshly passed faecal material from each major area in a given farm (for example, a pen). Culture of 136 pooled fecal samples from 29 livestock farms (Figure 4A) demonstrated that 6 (4%) samples from 4 (14%) farms (1 dairy cattle and 3 turkey) were positive for *K. pneumoniae* (Supplementary Table 1). Culture of 97 pre-packaged fresh meat products from 11 different countries purchased at 11 major Cambridge supermarkets were all negative for *K. pneumoniae* (Supplementary Table 2). Longitudinal sampling of sewage (4 samples over 16 months) at the study hospital (Figure 4A) resulted in *K. pneumoniae* being identified in 3/4 samples. A survey of 20 municipal wastewater treatment plants (Figure 4A) led to the recovery of *K. pneumoniae* from 17/40 water samples (11 untreated and 6 treated wastewater), taken from 12/20 treatment plants, with 5/12 plants releasing *K. pneumoniae* into the environment. Multiple colonies were selected from each positive sample for sequencing.

Genome-based comparison of K. pneumoniae from the hematology ward and elsewhere

We sequenced 87 *K. pneumoniae* isolates from livestock (32 isolates), municipal wastewater treatment plants (28 isolates) and hospital sewage (27 isolates). From this we identified four STs for livestock isolates, 24 STs from municipal wastewater and 6 STs from hospital sewage (Figure 2A). For

the livestock isolates, 2 STs (ST1609 and ST661) were identified from a single cattle farm, and one ST (ST3010) was isolated from all 3 turkey farms together with one colony of its single locus variant (SLV) from one of these (ST2585). All isolates from three different turkey farms (including the SLV) were a median (range) of 12 core genome SNPs (range 0-17) different, suggesting linkage between farms. Of these, two farms were located ~24 miles apart and the third was ~82 miles and 66 miles from the two other farms, respectively. All three farms were owned by the same company, and it is possible that transmission may have occurred as a result, although the route of transmission is unknown. STs were compared between isolates from the patient cohort, their ward environment, livestock, municipal and hospital waste (Figure 2A). There was no overlap in STs from the different sources with two exceptions (ST661 from patient stool & livestock, and ST20 from patient stool & municipal wastewater). The two ST611 isolates differed by 2641 core genome SNPs, and the 6 ST20 isolates differed by a median (range) of 849 (848-851) SNPs. This indicates that isolates from different sources that belonged to the same ST were not closely related.

A maximum-likelihood phylogenetic tree of the 249 study isolates (hematology ward (n=162), livestock/sewage/wastewater (n=87)) was constructed based on 499,378 core gene SNPs (Figure 2B). This demonstrated high genetic diversity and 3 distinct populations (208 KpI (*K. pneumoniae*), 22 KpII (*K*. *quasipneumoniae*) and 19 KpIII (*K*. *variicola*) isolates (Figure 2B), consistent with previous descriptions [1]. Each clade contained isolates from human stool and wastewater. Isolates from the hospital environment and blood were confined to KpI and KpIII, while isolates from cattle and turkeys resided in different clades (turkey in KpI, cattle in KpII).

Antibiotic resistance in K. pneumoniae from the hematology ward and elsewhere

Phenotypic antibiotic susceptibility of the 249 isolates to 19 antibiotics is summarized in Figure 4B. Resistance to meropenem and ertapenem was detected for two isolates from hospital sewage, both of which contained *bla*$_{KPC-2}$. No colistin resistance genes (MCR-1, MCR-1.2, MCR-2, MCR-3, MCR-4, MCR-5) were detected in any of the isolates. Colistin resistance in *K. pneumoniae* can also occur through modification of lipopolysaccharides that arise from mutations in the *pmrA*/*pmrB* and *phoP*/*phoQ* two-component regulatory systems. We identified the L96P mutation in *phoQ* associated previously with colistin resistance [24] in three isolates recovered from a single wastewater treatment plant (accession numbers: ERR985164, ERR985165, ERR985166). This indicates the importance of wastewater treatment plants as a source of resistance variants. Almost half of all isolates had an ESBL phenotype (118/249, 47%). Nearly all ESBL *K. pneumoniae* (n=113) carried *bla*$_{CTX-M-15}$, the remainder carrying *bla*$_{CTX-M-1}$ (n=3), or *bla*$_{KPC-2}$ and *bla*$_{SHV-12}$ (n=2). Isolates positive for *bla*$_{CTX-M-15}$ were distributed across 12 different STs and in all reservoirs tested, while *bla*$_{CTX-M-1}$ and *bla*$_{KPC-2}$ were reservoir specific and restricted to ST277 and ST258, respectively (Supplementary Table 1).

Plasmid analysis for the 113 *bla*$_{CTX-M-15}$ positive isolates was initially performed based on comparison between contigs containing the gene and the NCBI nucleotide database. This demonstrated close similarity to several reference plasmids, against which short read data for the 113 isolates were then mapped. All isolates from turkey farms contained a *bla*$_{CTX-M-15}$ plasmid with high sequence similarity (ID >99%, coverage >99%) to pKpN01-CTX [25]. Wastewater isolates that belonged to ST902 (municipal plant) and ST1436 (hospital sewage) showed high sequence match (ID >99%, coverage >98%) to pKPN3-307 Type A plasmid [26]. All ST268 isolates from stool and the ward environment mapped (ID >99%, coverage >99%) to the plasmid pCTXM15 (GenBank accession CP016925.1), consistent with these being associated with patient-environmental transmission. By contrast, whilst ST307 isolates from blood and stool mapped to IncFIB(K) pKPN3-307 Type B (GenBank: KY271405.1;

ID >97%, coverage >99%), ST307 isolates from the hospital environment were not positive for this plasmid and no plasmid reference was identified in the NCBI nucleotide database. This is consistent with these not being associated with patient-environment transmission. Long-read sequencing of one bloodstream and one environmental *bla*$_{\text{CTX-M-15}}$ ST307 isolate showed that both contained an IncFIB(K) plasmid (GenBank: MH745929 and MH745930, respectively). The environmental plasmid only had 78% coverage (ID >99%) against the human plasmid, however this shared sequence included a 31533-bp antimicrobial resistance region containing the *bla*$_{CTX\text{-}M\text{-}15}$ element (Figure 3B and Supplementary Table 3). Mapping of all ST307 *bla*$_{\text{CTX-M-15}}$ isolates to these two plasmids showed that clinical isolates contained MH745929 and the environmental isolates contained MH745930 (Figure 3A).

All *bla*$_{CTX\text{-}M\text{-}15}$ plasmids encoded a 13.4-kb region containing multiple antimicrobial resistance genes including *bla*$_{CTX\text{-}}$M-15, *bla*$_{TEM}$, *strA*, *strB* and *sul2* (see Figure 3B and AMR_Cluster1 in Supplementary Table 4). In total, 88/113 *bla*$_{CTX\text{-}M\text{-}15}$ positive isolates from humans, the hospital environment, livestock and wastewater had high coverage mapping (>99%) and sequence ID (>99%) to this cluster, which shared synteny with an antimicrobial resistance region found previously in other *K. pneumoniae* plasmids [26]. In addition, 17 isolates from the hospital sewer and humans shared a 7.2 kb resistance cluster (see AMR_Cluster2 in Supplementary Table 3), containing *bla*$_{CTX\text{-}M\text{-}15}$ and *aph(3)* (encoding aminoglycoside resistance).

Analysis of *bla*$_{CTX\text{-}M\text{-}1}$ and *bla*$_{\text{KPC-2}}$ positive isolates demonstrated that the three *bla*$_{CTX\text{-}M\text{-}1}$ isolates (ST277, from wastewater) mapped with high similarity to the p369 plasmid (GenBank accession number KT779550; ID >99%, coverage >99%) [27], and the two carbapenemase-producing *K. pneumoniae* (*bla*$_{\text{KPC-2}}$) isolates (ST258) from hospital sewage showed >99% mapping coverage and >99% sequence ID to the whole sequence of KPC reference plasmid pKpQIL-UK [28].

Virulence genes in K. pneumoniae from different reservoirs

The diversity of loci encoding capsule serotypes and the presence of specific virulence genes were identified in the 249 study isolates. Supplementary Table 1 provides a detailed description of capsule typing results. In brief, 43 different capsule loci were identified (Supplementary Fig. 1a), with 4 capsule loci with the same *wzi* gene allele were detected in more than one reservoir (Supplementary Fig. 1b). The *rmpA* gene (encoding a hypermucoid phenotype associated with increased capsule production) was identified in one isolate (Supplementary Table 1). Colibactin and aerobactin were only found in environmental isolates and salmochelin was not detected in any study isolate. Yersiniabactin and ICE*Kp* elements were identified in hospital sewage, wastewater, humans and the hospital environment, but ICE*Kp* types and yersiniabactin loci were predominantly reservoir and MLST-specific (Supplementary Table 1). ICE*Kp*10 was the only element that carried genes encoding yersiniabactin (*ybt*1), colibactin and aerobactin, and was present in ST23 isolates from the ward environment.

DISCUSSION

Here, we identified transmission of *K. pneumoniae* between patients and their ward environment and several broad environmental contamination events. A previous study suggested that *K. pneumoniae* is more transmissible than *E. coli* [29], but we did not capture episodes of patient-to-patient transmission. The extensive genetic diversity of *K. pneumoniae* isolated from patient stool is consistent with previous studies [14, 21]. Within-host diversity was limited, with 2/17 cases carrying more than one lineage and little diversity within the same lineage in a given host. We isolated *K. pneumoniae* from turkey, dairy cattle and wastewater, as described previously [30-34], but found no evidence to indicate that livestock or wastewater acted as a recent reservoir for *K. pneumoniae* isolated from patients. By contrast, we identified transmission of *K. pneumoniae* between farms.

Highly related plasmids carrying $bla_{CTX\text{-}M\text{-}15}$ were identified in isolates from the healthcare setting, and in isolates from non-healthcare settings, but we found no evidence for sharing of the same plasmid in healthcare and non-healthcare isolates. This suggests that livestock, wastewater and hospital sewage in this region were not a direct source of $bla_{CTX\text{-}M\text{-}15}$ plasmids in humans. Wastewater treatments were ineffective in eliminating *K. pneumoniae* from wastewater leading to downstream environmental contamination, which may contribute to the spread of antimicrobial resistance. Livestock isolates lacked virulence genes for yersiniabactin, salmochelin, aerobactin and colibactin, suggesting that livestock do not play a role in disseminating these virulence factors. Yersiniabactin was found in isolates from patients, municipal wastewater and hospital sewage, but the presence of an identical genetic locus (*ybt* 9) was only identified in ST268 isolates from a patient and their environment.

Our study has several limitations. We only recruited half of patients admitted to the two hematology wards, and alternative sources such as taps and patient food and water were not investigated. Overall, 17/149 (11%) patients were positive for stool carriage of *K. pneumoniae,* which is similar to that reported in Australia [21], but lower than a prevalence rate reported from the United States [15]. This low rate of recovery reduced the power to detect genetic relatedness between healthcare and non-healthcare associated isolates. Although we found no evidence for zoonotic transmission this may not reflect the situation elsewhere, particularly where people and livestock live in closer proximity.

In conclusion, our findings support the continued focus on reducing environmental *K. pneumoniae* reservoirs in hospital settings. Within the limits of detection of the study, our findings do not support

the suggestion that *K. pneumoniae* or their mobile genetic elements encoding antibiotic resistance are commonly acquired from livestock.

AUTHOR CONTRIBUTIONS

Study design: CL, TG and SJP. Developed study protocols: CL and TG. Collection of wastewater and farm samples: CL and TG. Sourcing and microbiological culture of retail meat samples: NH, TG and MH. Obtained access to farms and abattoirs: JHG, PW, MR and MH. Bacterial identification, susceptibility testing and phenotypic testing: CL, TG, BB and PN. Bioinformatics analyses and interpretation: CL, DM, DJ. Figure production: CL. Writing the manuscript: CL and SJP. Responsibility for supervision and management of the study: JP and SJP. All authors read and approved the final manuscript.

ACKNOWLEDGMENTS

The authors thank the Wellcome Trust Sanger Institute core library construction, sequence and informatics teams, and the Pathogen Informatics team. We thank the staff at farms, abattoirs and wastewater plants for assistance in sample collection, Elizabeth Lay for laboratory support during the meat survey. We thank Lois Chaparadza and Rosie Swayne for support during the ward study. We also thank Alexander M. Wailan for advice on plasmid biology. The flocked swabs used in the meat survey were donated by Copan Italia spa. The study protocol was approved by the Cambridge University Hospitals NHS Foundation Trust Research and Development Department (ref: A093285), the Department of Veterinary Medicine, University of Cambridge (study CR112) and the National Research Ethics Service East of England Ethics Committee (ref: 12/EE/0439 and 14/EE/1123). Sequence data for all isolates have been submitted to the European Nucleotide Archive (www.ebi.ac.uk/ena) under study accession number PRJEB18814 and PRJEB19918 (all human, environmental and hospital sewage *K. pneumoniae*), PRJEB8771 and PRJEB8776 (*K. pneumoniae* from livestock), PRJEB8766 and PRJEB8770 (*K. pneumoniae* from wastewater) with the accession numbers for individual isolates listed in Supplementary Table 1.

FUNDING

This work was supported by the Health Innovation Challenge Fund (WT098600, HICF-T5-342), a parallel funding partnership between the Department of Health and Wellcome Trust. The views expressed in this publication are those of the author(s) and not necessarily those of the Department of Health or Wellcome Trust. This project was also funded by a grant awarded to the Wellcome Trust Sanger Institute (098051). TG is a Wellcome Trust Research Training Fellow (103387/Z/13/Z). CL is a Wellcome Trust Sir Henry Postdoctoral Fellow (110243/Z/15/Z). DJ is funded by the Wellcome Trust grant 098051.

POTENTIAL CONFLICT OF INTEREST

JP reports personal fees from Next Gen Diagnostics, outside the submitted work.

REFERENCES

1. Holt KE, Wertheim H, Zadoks RN, 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.

2. Podschun R, Ullmann U. *Klebsiella* spp. as nosocomial pathogens: epidemiology, taxonomy, typing methods, and pathogenicity factors. Clinical Microbiology Reviews **1998**; 11(4): 589-603.

3. Wyres KL, Holt KE. *Klebsiella pneumoniae* population genomics and antimicrobial-resistant clones. Trends in Microbiology **2016**; 24(12): 944-56.

4. Snitkin ES, Zelazny AM, Thomas PJ, et al. Tracking a hospital outbreak of carbapenem-resistant *Klebsiella pneumoniae* with whole-genome sequencing. Sci Transl Med **2012**; 4.

5. Clarivet B, Grau D, Jumas-Bilak E J, et al. Persisting transmission of carbapenemase-producing *Klebsiella pneumoniae* due to an environmental reservoir in a university hospital, France, 2012 to 2014. Euro Surveill **2016**; 21(17): 30213.

6. Marsh JW, Krauland MG, Nelson JS, et al. Genomic epidemiology of an endoscope-associated outbreak of *Klebsiella pneumoniae* carbapenemase (KPC)-producing *K. pneumoniae*. PLOS ONE **2015**; 10(12): e0144310.

7. Brooks B, Olm MR, Firek BA, et al. Strain-resolved analysis of hospital rooms and infants reveals overlap between the human and room microbiome. Nature Communications **2017**; 8(1): 1814.

8. Overdevest I, Willemsen I, Rijnsburger M, et al. Extended-spectrum β-lactamase genes of *Escherichia coli* in chicken meat and humans, the Netherlands. Emerging Infectious Diseases **2011**; 17(7): 1216-22.

9. Davis GS, Waits K, Nordstrom L, et al. Intermingled *Klebsiella pneumoniae* populations between retail meats and human urinary tract infections. Clinical Infectious Diseases **2015**; 61(6): 892-9.

10. Kim S-H, Wei C-I, Tzou Y-M, An H. Multidrug-resistant *Klebsiella pneumoniae* isolated from farm environments and retail products in Oklahoma. Journal of Food Protection **2005**; 68(10): 2022-9.

11. Munoz MA, Ahlström C, Rauch BJ, Zadoks RN. Fecal Shedding of *Klebsiella pneumoniae* by Dairy Cows. Journal of Dairy Science **2006**; 89(9): 3425-30.

12. Galler H, Feierl G, Petternel C, et al. KPC-2 and OXA-48 carbapenemase-harbouring Enterobacteriaceae detected in an Austrian wastewater treatment plant. Clinical Microbiology and Infection **2014**; 20(2): O132-O4.

13. Runcharoen C, Moradigaravand D, Blane B, et al. Whole genome sequencing reveals high-resolution epidemiological links between clinical and environmental *Klebsiella pneumoniae*. Genome Medicine **2017**; 9(1): 6.

14. Gorrie CL, Mirčeta M, Wick RR, et al. Gastrointestinal carriage is a major reservoir of *Klebsiella pneumoniae* infection in intensive care patients. Clinical Infectious Diseases **2017**; 65(2): 208-15.

15. Martin RM, Cao J, Brisse S, et al. Molecular epidemiology of colonizing and infecting isolates of *Klebsiella pneumoniae*. mSphere **2016**; 1(5).

16. Page AJ, De Silva N, Hunt M, et al. Robust high-throughput prokaryote de novo assembly and improvement pipeline for Illumina data. Microbial Genomics **2016**; 2(8).

17. Zerbino DR, Birney E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res **2008**; 18.

18. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics **2014**; 30(14): 2068-9.

19. Croucher NJ, Page AJ, Connor TR, et al. Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. Nucleic Acids Research **2015**; 43(3): e15.

20. Letunic I, Bork P. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic Acids Research **2016**.

21. Gorrie CL, Mirceta M, Wick RR, et al. Antimicrobial resistant *Klebsiella pneumoniae* carriage and infection in specialized geriatric care wards linked to acquisition in the referring hospital. Clinical Infectious Diseases **2018**: ciy027-ciy.

22. Inverarity D, Kilgour E, Dunn C, et al. Screening haematology patients for carbapenem-resistant *Klebsiella pneumoniae*. Journal of Infection Prevention **2014**; 15(2): 50-6.

23. Freeman JT, Nimmo J, Gregory E, et al. Predictors of hospital surface contamination with extended-spectrum β-lactamase-producing *Escherichia coli* and *Klebsiella pneumoniae*: patient and organism factors. Antimicrobial Resistance and Infection Control **2014**; 3: 5-.

24. Olaitan AO, Diene SM, Kempf M, et al. Worldwide emergence of colistin resistance in *Klebsiella pneumoniae* from healthy humans and patients in Lao PDR, Thailand, Israel, Nigeria and France owing to inactivation of the PhoP/PhoQ regulator mgrB: an epidemiological and molecular study. International Journal of Antimicrobial Agents **2014**; 44(6): 500-7.

25. Lynch T, Chen L, Peirano G, et al. Molecular evolution of Klebsiella pneumoniae ST278 harboring blaNDM-7 involved in nosocomial transmission. Journal of Infectious Diseases **2016**: jiw240.

26. Villa L, Feudi C, Fortini D, et al. Diversity, virulence, and antimicrobial resistance of the KPC-producing *Klebsiella pneumoniae* ST307 clone. Microbial Genomics **2017**; 3(4): e000110.

27. Baron S, Jouy E, Touzain F, et al. Impact of the administration of a third-generation cephalosporin (3GC) to one-day-old chicks on the persistence of 3GC-resistant *Escherichia coli* in intestinal flora: an in vivo experiment. Veterinary Microbiology **2016**; 185(Supplement C): 29-33.

28. Doumith M, Findlay J, Hirani H, et al. Major role of pKpQIL-like plasmids in the early dissemination of KPC-type carbapenemases in the UK. Journal of Antimicrobial Chemotherapy **2017**; 72(8): 2241-8.

29. Gurieva T, Dautzenberg MJD, Gniadkowski M, Derde LPG, Bonten MJM, Bootsma MCJ. The transmissibility of antibiotic-resistant Enterobacteriaceae in intensive care units. Clinical Infectious Diseases **2018**; 66(4): 489-93.

30. Kim S-H, Wei C-I, Tzou Y-M, An H. Multidrug-Resistant Klebsiella pneumoniae Isolated from Farm Environments and Retail Products in Oklahoma, **2005**.

31. He T, Wang Y, Sun L, Pang M, Zhang L, Wang R. Occurrence and characterization of blaNDM-5-positive *Klebsiella pneumoniae* isolates from dairy cows in Jiangsu, China. Journal of Antimicrobial Chemotherapy **2017**; 72(1): 90-4.

32. Munoz MA, Welcome FL, Schukken YH, Zadoks RN. Molecular epidemiology of two *Klebsiella pneumoniae* mastitis outbreaks on a dairy farm in New York state. Journal of Clinical Microbiology **2007**; 45(12): 3964-71.

33. Dolejska M, Frolkova P, Florek M, et al. CTX-M-15-producing *Escherichia coli* clone B2-O25b-ST131 and *Klebsiella* spp. isolates in municipal wastewater treatment plant effluents. Journal of Antimicrobial Chemotherapy **2011**; 66(12): 2784-90.

34. Chagas TPG, Seki LM, da Silva DM, Asensi MD. Occurrence of KPC-2-producing *Klebsiella pneumoniae* strains in hospital wastewater. Journal of Hospital Infection **2011**; 77(3): 281.

FIGURE LEGENDS

Figure 1. Patient recruitment and phylogeny of healthcare-associated *K. pneumoniae* isolates

(A) Patients recruited and samples collected from study participants. (B) Maximum likelihood tree of isolates from patient stool, blood and the hospital environment based on SNPs in the core genome.

Figure 2. Relatedness of *K. pneumoniae* isolated from different sources

(A) Distribution of multilocus sequence types (STs) by source of isolation. (B) Maximum-likelihood core genome phylogeny of *K. pneumoniae* from humans, the hospital environment, hospital sewage, livestock and municipal wastewater from the East of England. The three clades are highlighted in orange (KpI), blue (KpII) and green (KpIII). The two right hand columns (from left to right) show isolate source, and those cases where STs were the same for both human and non-human isolates.

Figure 3. Phylogeny of ST307 *K. pneumoniae* isolates and characterisation of plasmids present in isolates from patients and their environment

(A) Maximum likelihood tree of ST307 *K. pneumoniae* isolates from patient stool, blood and the hospital environment based on SNPs in the core genome after removal of recombination. (B) Antimicrobial resistance genes and associated mobile elements in the pKPN3-307 Type B (GenBank: KY271405.1) reference plasmid and in the human (GenBank: MH745929) and environmental plasmids (GenBank: MH745930). The shared 31533-bp antimicrobial resistance region containing the $bla_{CTX\text{-}M\text{-}15}$ element is highlighted in grey. Arrows indicate the orientation of features, with the forward direction defined as the direction of transcription for genes, towards the main part of the attC site for cassettes, in integrons towards attI for 5' flanking regions, away from the cassette array for 3'-flanking regions, relative to the direction of transcription of the transposase gene for IS and Tn

(i.e. IRL to IRR) and to the direction of the reverse transcriptase for Group II introns. The missing end of a feature is shown by a zig-zag line.

Figure. 4. Sampling locations and phenotypic antimicrobial susceptibility across the *K. pneumoniae* phylogeny

(A) Map of the East Anglia region of the UK, showing the locations of the farms (images indicate the animal species), wastewater treatment plants (water drops), and hospitals in the region (indicated by a white 'H' in a blue square). The hospital where the clinical and environmental isolates were collected is shown with H surrounded by a square. Adapted from "Genomic surveillance of *Enterococcus faecium* reveals limited sharing of strains and resistance genes between livestock and humans in the United Kingdom," by Gouliouris et al, 2018, *mBio,* 9 (6) e01780-18. Copyright 2018 Gouliouris et al. Adapted with permission. (B) Maximum likelihood core genome phylogeny of 249 *K. pneumoniae* from human stool, blood, livestock, municipal wastewater and hospital sewage (left), and their phenotypic antimicrobial susceptibility (right). The first vertical column shows isolate source and the remainder the susceptibility to 19 antimicrobial drugs.

Figure 1

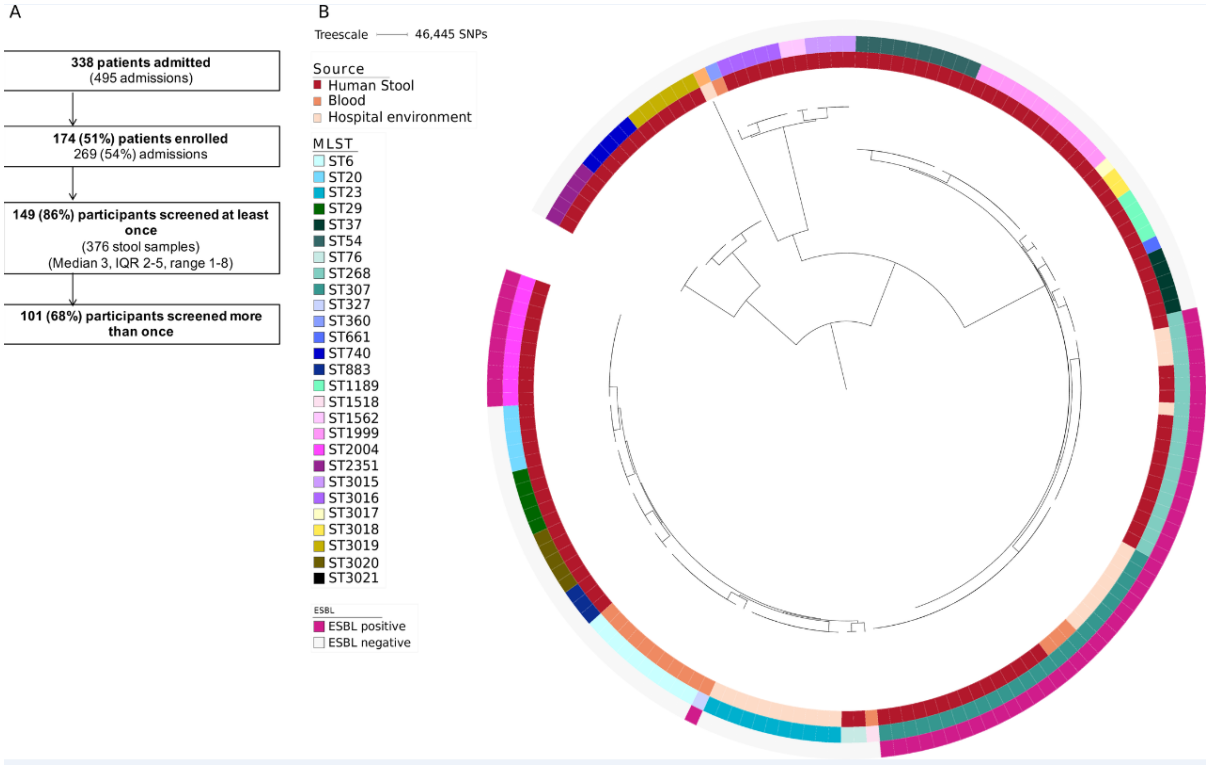

Figure 2

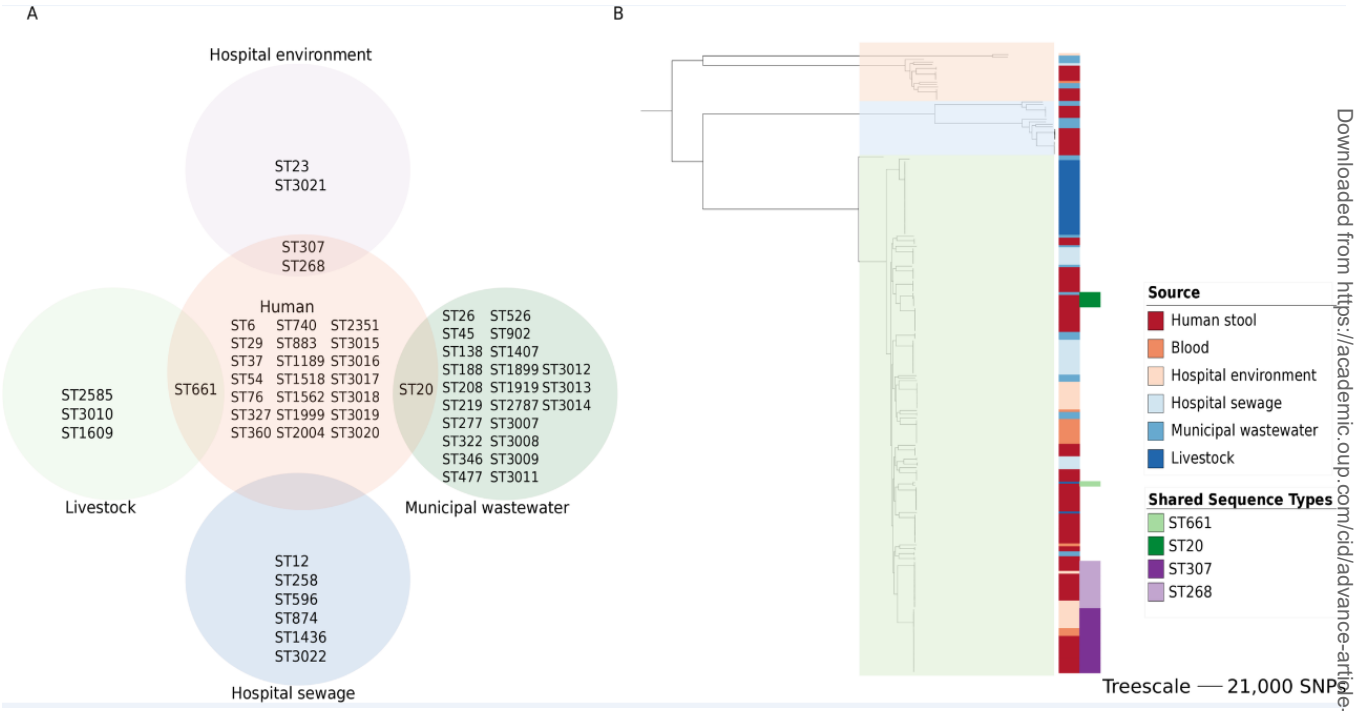

Figure 3

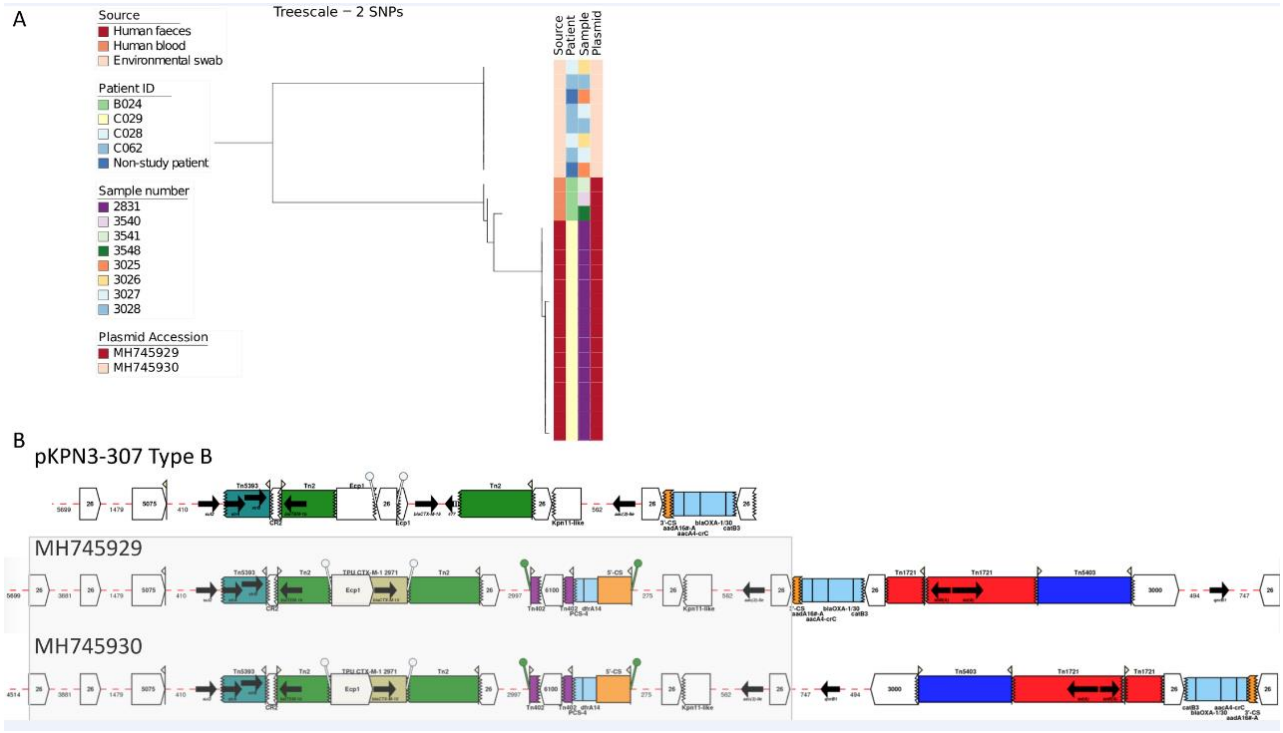

Figure 4