

RESEARCH

Open Access



Whole genome sequencing reveals high-resolution epidemiological links between clinical and environmental *Klebsiella pneumoniae*

Chakkaphan Runcharoen¹, Danesh Moradigaravand^{2*}, Beth Blane³, Suporn Paksanont¹, Jeeranan Thammachote⁴, Suthatip Anun⁴, Julian Parkhill², Narisara Chantratita^{1,5} and Sharon J. Peacock^{2,3,6*}

Abstract

Background: *Klebsiella pneumoniae* is a gram-negative bacterial species capable of occupying a broad range of environmental and clinical habitats. Known as an opportunistic pathogen, it has recently become a major causative agent of clinical infections worldwide. Despite growing knowledge about the highly diverse population of *K. pneumoniae*, the evolution and clinical significance of environmental *K. pneumoniae*, as well as the relationship between clinical and environmental *K. pneumoniae*, are poorly defined.

Methods: We isolated and sequenced *K. pneumoniae* from in-patients in a single hospital in Thailand, as well as hospital sewage, and surrounding canals and farms within a 20-km radius.

Results: Phylogenetic analysis of 77 *K. pneumoniae* (48 clinical and 29 non-clinical isolates) demonstrated that the two groups were intermixed throughout the tree and in some cases resided in the same clade, suggesting recent divergence from a common ancestor. Phylogenetic comparison of the 77 Thai genomes with 286 *K. pneumoniae* from a global collection showed that Thai isolates were closely related to the clinical sub-population of the global collection, indicating that Thai clinical isolates belonged to globally circulating lineages. Dating of four Thai *K. pneumoniae* clades indicated that they emerged between 50 and 150 years ago. Despite their phylogenetic relatedness, virulence factors and β -lactamase resistance genes were more numerous in clinical than in environmental isolates. Our results indicate that clinical and environmental *K. pneumoniae* are closely related, but that hospitals may select for isolates with a more resistant and virulent genotype.

Conclusions: These findings highlight the clinical relevance of environmental *K. pneumoniae* isolates.

Keywords: *Klebsiella pneumoniae*, Whole genome sequencing, Antibiotic resistance, Genomic epidemiology

Background

Klebsiella pneumoniae is a clinically important gram-negative bacterium associated with opportunistic infection in patients with a compromised immune system or receiving other forms of complex medical care [1–3]. This species is disseminated in healthcare settings via person-to-person contact, medical equipment, and contaminated

environments [2, 3]. The emergence of multidrug-resistant *K. pneumoniae* is becoming an increasingly serious issue for clinical practice, largely related at the present time to isolates that express extended-spectrum β -lactamase (ESBL) enzymes that hydrolyze a broad spectrum of β -lactams [4, 5]. This is compounded by the emergence of *K. pneumoniae* that express carbapenemases such as KPC-type β -lactamase [5–7] and the recent detection of colistin-resistant *K. pneumoniae* due to the presence of the *mcr-1* gene [8]. These resistance genes are carried on mobile genetic elements that facilitate their spread within and between bacterial

* Correspondence: dm16@sanger.ac.uk; sjp97@medschl.cam.ac.uk
Chakkaphan Runcharoen and Danesh Moradigaravand are joint first author. Narisara Chantratita and Sharon J. Peacock are joint senior authors.

²Wellcome Trust Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridgeshire, UK

Full list of author information is available at the end of the article

species, a process that is likely to result in a rise in the number of *K. pneumoniae* infections that are very difficult to treat. Furthermore, multidrug-resistant *K. pneumoniae* are associated with nosocomial outbreaks, particularly in high prevalence countries including those in East Asia [9, 10].

Beyond healthcare settings and hospital patients, *K. pneumoniae* is ubiquitous in nature and occupies a diverse range of niches. These include environmental sources, such as soil and wastewater, mucosal surfaces and the gut of humans and animals, and food sources, such as meat [11]. Environmental *K. pneumoniae* is less well studied than isolate collections associated with clinical disease. Some studies have shown that *K. pneumoniae* of environmental origin are highly similar to clinical isolates with regards to phenotypic and some genetic features [12–14], but others have reported differences in virulence characteristics between the two groups [15, 16]. The parallel evolution of *K. pneumoniae* and putative acquisition of antimicrobial resistance determinants and virulence factors in healthcare settings and the environment have led non-clinical habitats to be considered as potential reservoirs for hyper-virulent and hyper-resistant *K. pneumoniae* [15], although evidence to support the potential clinical importance of non-clinical *K. pneumoniae* is inconclusive. A recent study of the clinical relevance of meat-source *K. pneumoniae* showed differences in antibiotic resistance but similar virulence characteristics for isolates from retail meat and those associated with urinary tract infections in humans [16].

Here, we report the findings of an in-depth comparison of the evolution and epidemiology of ESBL-positive *K. pneumoniae* using a One Health approach [17, 18]. We utilized the fine-scale resolution of whole genome sequencing to investigate genetic relatedness, antimicrobial resistance, and the presence of genes encoding virulence determinants in an unbiased, prospective collection of *K. pneumoniae* from patients in one hospital, as well as from environmental water, hospital sewage, and farm waste in the proximity of the hospital. These genomes were placed into a global context through a comparison with isolates recovered from various clinical and non-clinical sources worldwide [19]. Our results highlight the clinical relevance of environmental *K. pneumoniae* isolates, and demonstrate that environmental and clinical *K. pneumoniae* are highly related but that hospitals select for *K. pneumoniae* with a more antimicrobial-resistant and virulent genotype.

Methods

Study design and bacterial isolates

The bacterial collection consisted of 77 ESBL-producing *K. pneumoniae* isolated between 2014 and 2015. Clinical isolates ($n = 48$) were obtained from consecutive patients

with positive samples processed by the diagnostic microbiology laboratory at Bhuddhasothorn Hospital, Chachoengsao, Bangkok, Thailand between December 2014 and April 2015. Data were collected on date of isolation and sample type, and samples were de-duplicated so that only one isolate per patient was included. Speciation and ESBL positivity were initially determined using Standard Operating Procedures supplied by the Department of Medical Science, Ministry of Public Health, Thailand and Clinical and Laboratory Standards Institute (CLSI) guidelines (M100-S24 and M100-S25), respectively. The species was subsequently confirmed using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS; Biotyper version 3.1, Bruker Daltonics, Coventry, UK). Antimicrobial susceptibility testing was repeated using the N206 card on the Vitek 2 instrument (bioMérieux, Marcy l'Étoile, France) calibrated against EUCAST breakpoints, and these results were used during the analysis.

Environmental and livestock-associated isolates ($n = 29$) were obtained through a cross-sectional survey between January 2015 and February 2015. Wastewater samples were collected from 27 canals and 11 farms within a 20-km radius of Bhuddhasothorn Hospital. The farms reared pigs ($n = 2$), chickens ($n = 6$), ducks ($n = 2$), and both chickens and ducks ($n = 1$). Samples were collected from wastewater collection areas (commonly concrete gullies that drained waste from animal housing). The longitude and latitude of each sampling site were recorded using GPSMAP 60CSx (Garmin, Taiwan). A further two wastewater samples were taken from the Bhuddhasothorn Hospital wastewater treatment system (one pre-treatment and one post-treatment water sample). At each site, grab samples of 0.5 L each were collected into sterile bottles containing 9 mg of sodium thiosulfate pentahydrate (Merck, Darmstadt, Germany). All samples were transported to the laboratory on ice packs in the dark and processed within 12 h.

One mL of triplicate serial 10-fold dilutions of wastewater samples were concentrated using the filtration technique onto 0.45- μ m pore size filter membranes (Merck, Darmstadt, Germany). Membranes were then placed onto the surface of ESBL Brilliance agar (Oxoid, Basingstoke, UK) and incubated for 48 h at 35 °C in air. For each sampling site, up to 10 colonies suspected to be *K. pneumoniae* based on color (green) were picked and the presence of ESBL confirmed using the combination disc test (M100-S24 and M100-S25). *Escherichia coli* ATCC25922 and *K. pneumoniae* ATCC700603 were used as negative and positive controls, respectively. Positive colonies were speciated using MALDI-TOF MS, and antimicrobial susceptibility testing of confirmed *K. pneumoniae* was determined using the N206 card as described above. All isolates were stored at -80 °C until further analysis.

Whole genome sequencing and pan-genome analysis

DNA extraction, sequencing, and assembly of reads were performed as previously described [20]. Sequencing was performed on an Illumina HiSeq2000. Details of reads, depth of coverage, and N50 are provided in Additional file 1. The sequence reads were submitted to the European Nucleotide Archive (ENA) under accession numbers [ENA:ERP012787 and ENA:PRJEB11403]. An average coverage of 85-fold was achieved. Genomes were assembled using Velvet [21] with the pipeline and improvements found at <https://github.com/sanger-pathogens/vr-codebase> and https://github.com/sanger-pathogens/assembly_improvement. The de novo assemblies were annotated using Prokka [22], and sequence types (STs) were identified from the sequence data using a multilocus sequence typing (MLST) database (<http://bigsd.b.pasteur.fr/klebsiella/>) and an in-house script (https://github.com/sanger-pathogens/mlst_check).

Phylogenetic analysis and substitution rate calculation

Study genomes were contextualized against a global collection. Sequence data for 286 *K. pneumoniae* isolates reported previously [19] were downloaded from the ENA and combined with the 77 study genomes. As in our study, these isolates were also sequenced on a HiSeq sequencing system. Short reads for the 363 isolates were mapped against the reference genome *K. pneumoniae* Ecl8 (accession number: HF536482 CANH01000000) using SMALT v0.7.4 (<http://www.sanger.ac.uk/science/tools/smalt-0>). An in-house tool that combined SAMtools mpileup [23] and BCFtools, as detailed in [24], was used to annotate single nucleotide polymorphisms (SNPs), after which the pairwise SNP distances were calculated from the multiple alignment to obtain the phylogenetic tree. Mapping each genome to the reference genome allowed us to identify genes that were conserved in the core genome of the study isolates and reference genome. We constructed a maximum likelihood tree using FastTree version 2.1.3 with 100 bootstraps and a midpoint root [25]. We employed FigTree (<http://tree.bio.ed.ac.uk/software/figtree/>), Microreact (www.microreact.org), and in-house tools to visualize the results.

To determine the substitution rate for each clade (see the Results section for more details), reads were first mapped within each clade to the sequence obtained after concatenating contigs for the isolate with the best contigs statistics, i.e., the isolate with the highest N50 value. After mapping the reads to the new references, high-density SNP regions (which are indicative of putative recombination events) were removed from the multiple alignments using Gubbins, which works best for closely related isolates and detects high SNP density regions based on a significantly higher number of variable sites in a sliding window across the genome compared to the

rest of the genome [26]. The significance of the temporal signal was assessed according the *R*-squared value obtained from the root-to-tip distance versus time of isolation plot. We generated 10,000 sample sets by bootstrapping and assessed the value of *R*-squared against the distribution of *R*-squared values. Of four clades we identified on the phylogenetic tree of the Thai and global isolates, we found strong signals (>99%) for clades 1 and 3, after leaving one isolate out from either clade. BEAST version 1.7 was used to estimate the substitution rate and date the phylogenetic tree for the clades with significant temporal signal [27], using a strict molecular clock and a lognormal and uniform prior distribution models for base frequencies with constant population size. Three chains of BEAST were run for 50 million generations (sampling every 10 generations) and checked as to whether the runs had converged on similar values. Convergence was controlled using effective sample size (ESS) value (we considered a cut-off of 200 ESS for convergence). We excluded 10 million initial steps as a burn-in phase and merged the out trees with the Tree-Annotator program in the BEAST package, and chose the model in which convergence always occurred (for the clades we tested here, the strict molecular clock with a uniform prior distribution for base frequencies always converged).

Identification of antimicrobial resistance determinants, virulence factors, and plasmids

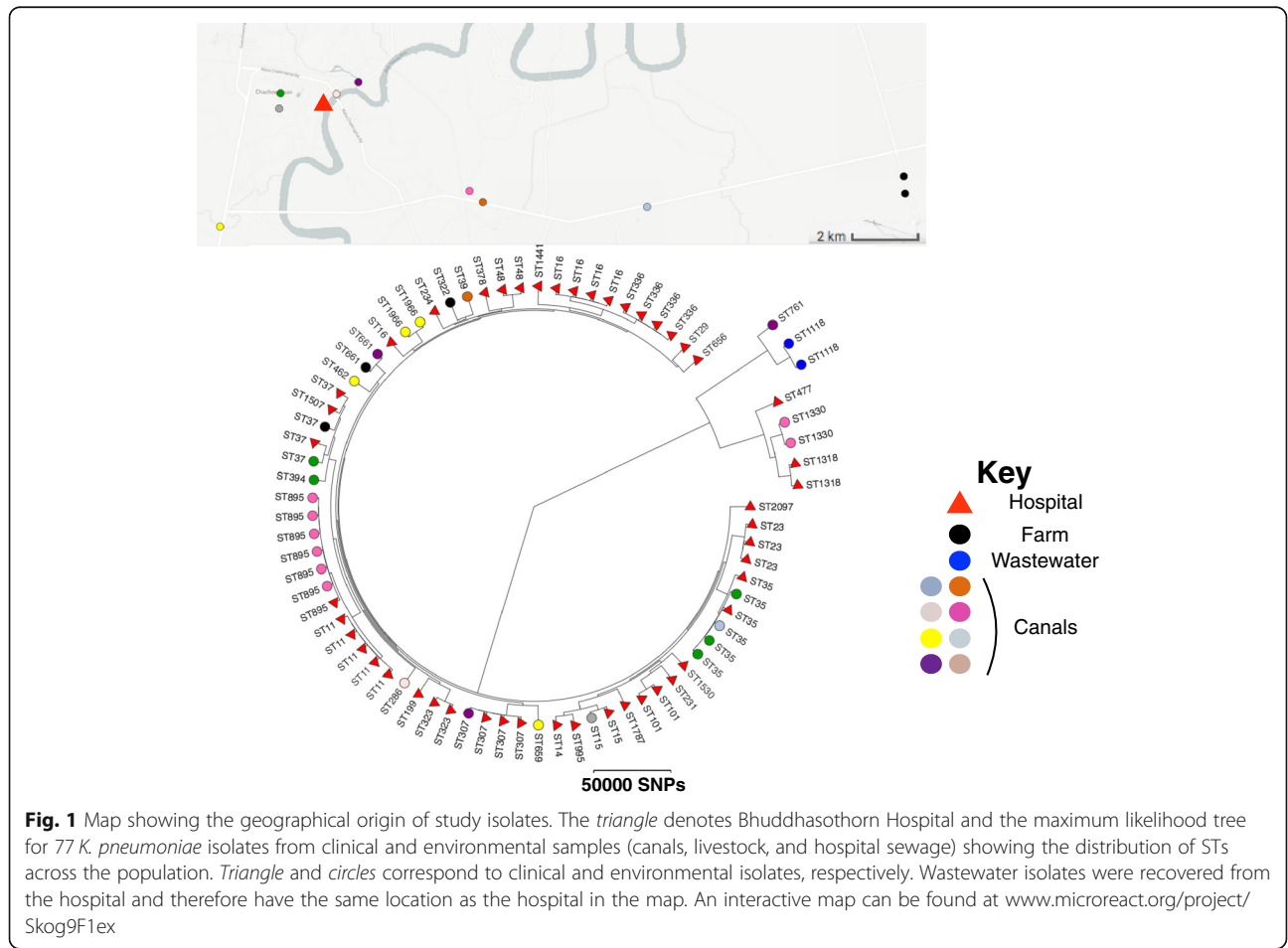
We employed the srst2 package [28], which takes short reads and maps them to reference sequences, to find antibiotic resistance genes, virulence factors, and plasmid replicons. The sequences for 79 virulence genes were obtained from the Pasteur institute data repository (<http://bigsd.b.pasteur.fr/klebsiella/klebsiella.html>, http://bigsd.b.pasteur.fr/perl/bigsd/bigsd.pl?db=pubmlst_klebsiella_seqdef_public&page=downloadAlleles). An in-house tool was used to visualize the results. To find the possible context of the virulence (i.e., chromosomal or plasmid based), we first found the contig in which the virulence factor was located and then extracted the 5-kb sequences upstream and downstream of the gene; if this exceeded the contig size, we considered the end and start of the contig. We then performed blast searches using the NCBI non-redundant nucleotide database to find out whether the hit sequences corresponded to a chromosomal or a plasmid region. Furthermore, we assessed the significance between the presence of virulence factors in the clinical/environmental sub-populations by performing logistic regression. To this end, we took the presence/absence of individual genes as the categorical predictor parameter and the environmental/clinical status as the categorical dependent variable. Significance of association (*p* value < 0.05) was then assessed by considering the *z*-statistic, which is the regression coefficient divided by its standard error and has a standard normal distribution.

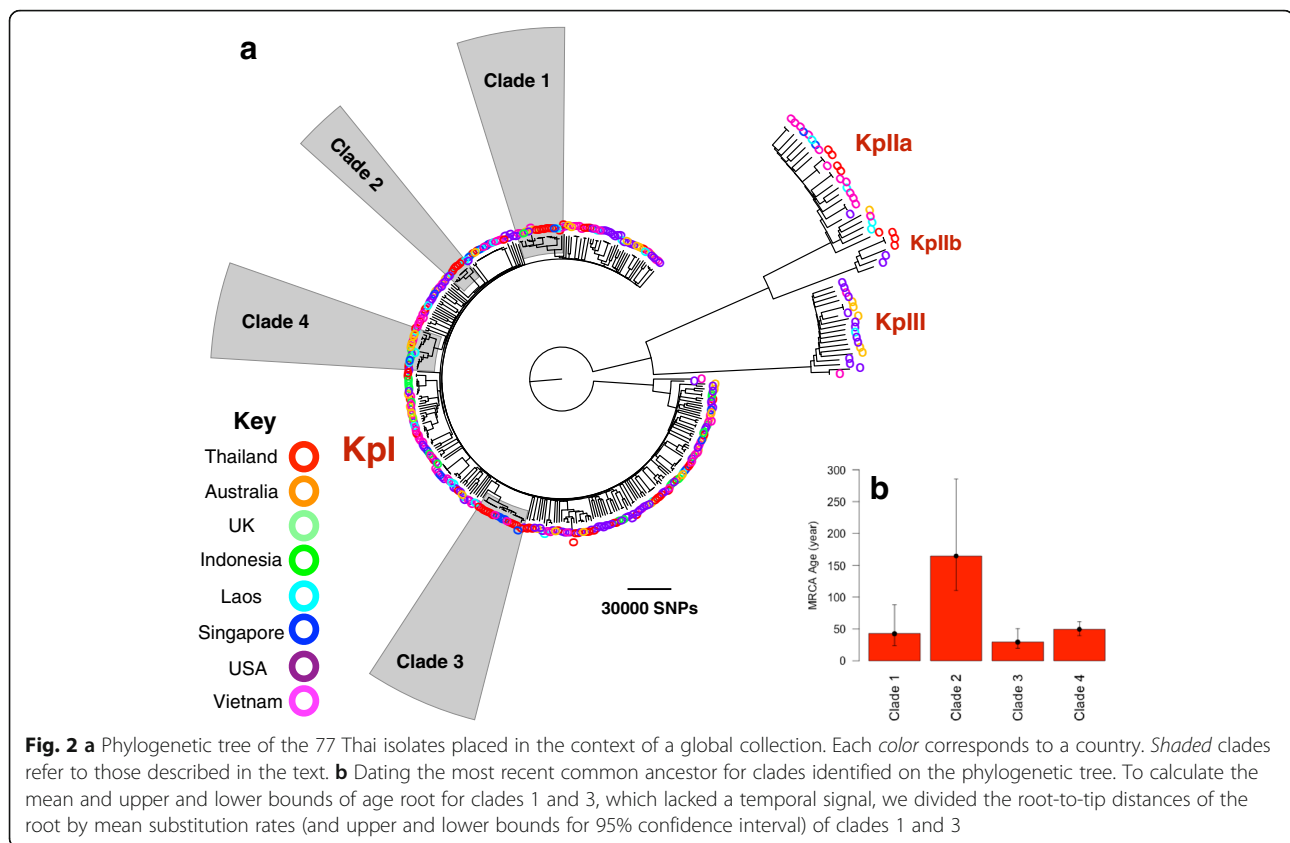
Results

A One Health approach was taken for the sampling and isolation of ESBL-positive *K. pneumoniae* from clinical samples in a hospital in Thailand, together with hospital sewage, environmental (canal) water, and wastewater from farms within a 20-km radius of the hospital. A phylogenetic tree of the whole genomes of the 77 *K. pneumoniae* isolates (48 clinical and 29 environmental, of which 24 were from canals, 3 from farms, and 2 from untreated hospital sewage) revealed a diverse population containing three major lineages, one of which contained the majority of isolates (Fig. 1). Several minor clades were also apparent, some of which showed evidence of recent expansion (Fig. 1). Clinical and environmental isolates were intermixed throughout the tree and in some cases resided in the same clade, suggesting recent divergence from a common ancestor (Fig. 1). The 3 isolates of farm origin were distributed across the tree, and clustered with canal and clinical isolates or with canal isolates alone (Fig. 1). High genetic diversity was also reflected by the number of STs ($n = 38$) to which the 77 isolates were assigned. This included STs that have been isolated elsewhere in the world. For example, ST35,

ST307, and ST16 were recently recovered from clinical settings in France, the USA, and the Netherlands, respectively [29–31]. Mapping the geographical distribution of isolates recovered from the environment demonstrated that isolates from the same sampling site were frequently very genetically diverse (Fig. 1).

To further characterize the 77 study genomes, we combined them with genomes for a global collection of 286 *K. pneumoniae complex* isolates [19]. The resulting phylogenetic tree revealed that Thai isolates were dispersed across the combined population. Isolates in the global collection had been defined previously based on phylogroup, and Thai isolates were observed to cluster in phylogroups KpI, KpIIa, and KpIIb, with the majority residing in KpI (Fig. 2a). This broader genetic context highlighted the presence of several clades of Thai isolates; those containing at least 4 Thai isolates (clades 1 to 4) were subjected to more detailed analysis to uncover their recent evolutionary history. These clades included 30 of the 77 total isolates. A strong temporal signal was found for clade 1 and clade 3 after removing regions of recombination. This allowed us to estimate the substitution rates, which were 8.33×10^{-7} (95% confidence



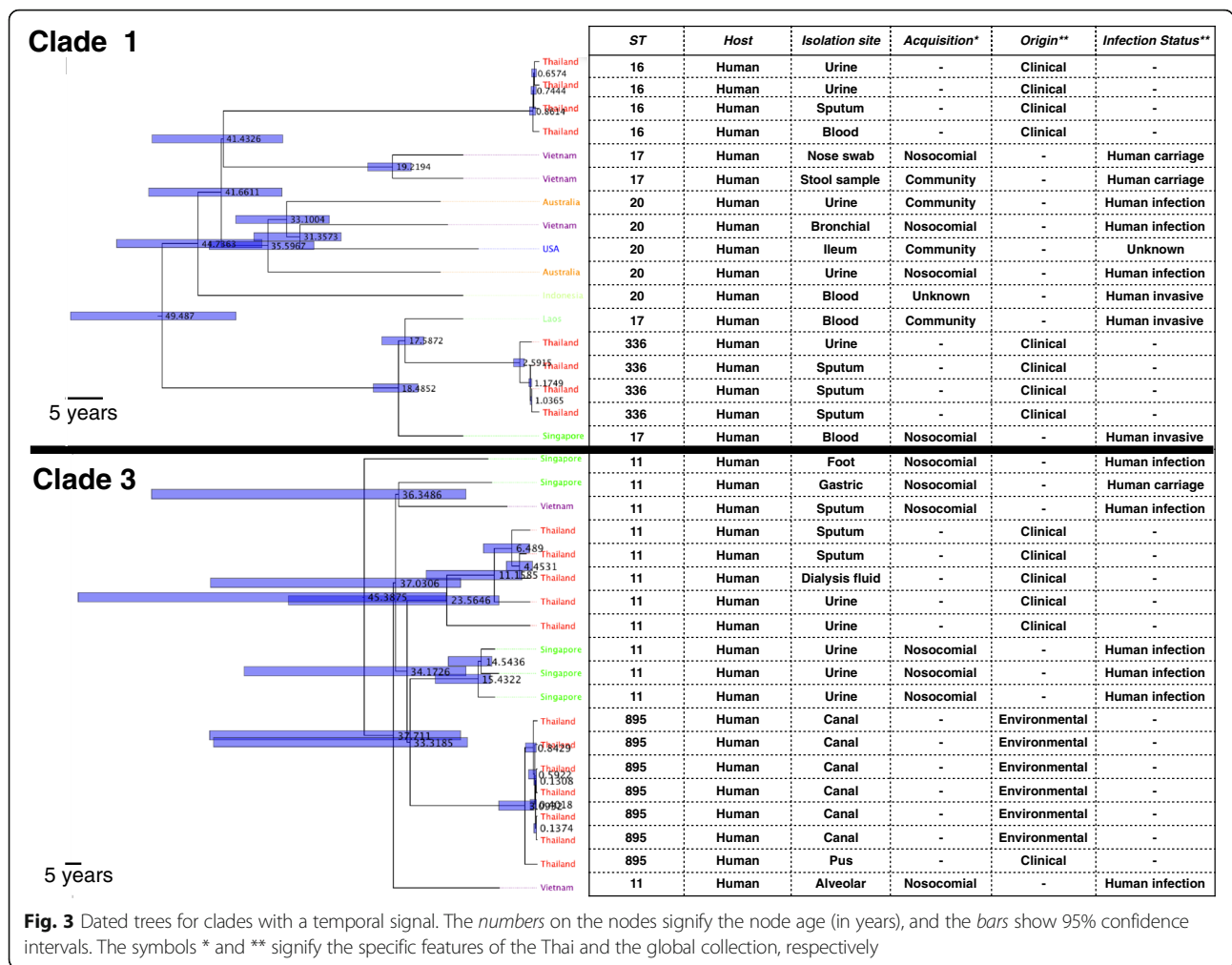


interval [CI]: 6.23×10^{-7} , 1.05×10^{-6}) per site per year for clade 1 and 3.78×10^{-7} (95% CI: 7.47×10^{-8} , 7.54×10^{-7}) per site per year for clade 3. The most recent common ancestor (MRCA) for clades 1 and 3 was estimated to exist 50 to 70 years ago (Fig. 2b). Using the average substitution rates for clades 1 and 3, we also estimated the age of the MRCA for clades 2 and 4 to be 150 and 50 years ago, respectively (Fig. 2b). We conclude that expansion of these *K. pneumoniae* clones has taken place over the past few decades.

Thai isolates residing in clades 1 and 3 were intermixed with isolates from other countries in the global collection (Fig. 3). The Thai isolates in clade 1 fell into two sub-clades of clinical origin, both of which were estimated to have emerged in the past few decades. These were more distantly related to two clinical isolates from community and nosocomial infection in the global collection (Fig. 3). Clade 3 also contained two Thai sub-clades, one of which was composed of clinical isolates. The second sub-clade consisted largely of environmental isolates but had recently diverged from a clinical Thai isolate (Fig. 3). Global isolates in clade 3 were of hospital origin and nosocomial infections.

To generalize our findings about the epidemiological links between global and clinical and environmental Thai isolates for the isolates not included in clades 1 or 3, we

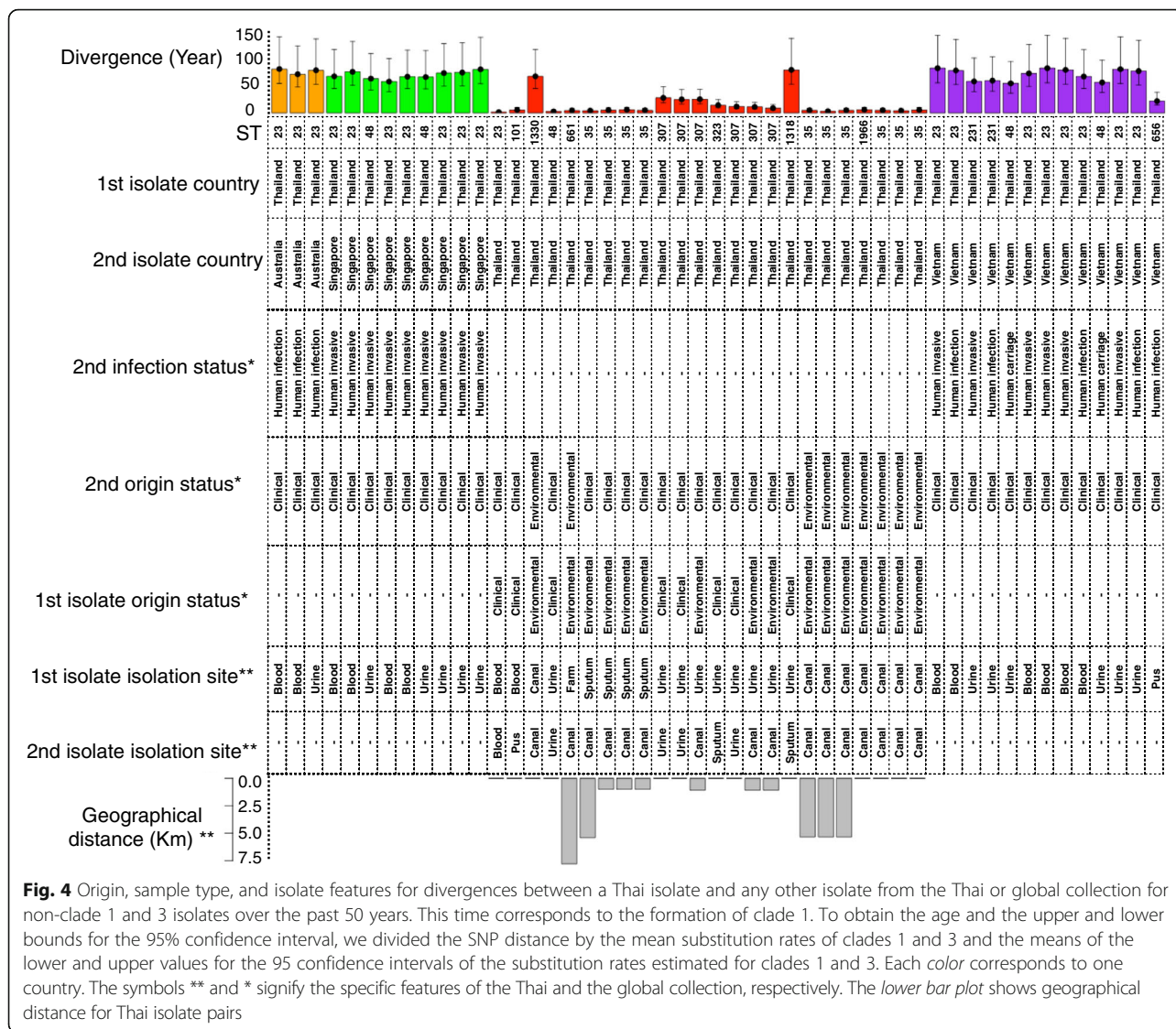
excluded the isolates in the two clades and then used the average substitution rates for isolates in clades 1 and 3 to estimate divergence times between a Thai isolate and any other Thai or global isolate that was less than 250 SNPs apart (this corresponds to approximately 50 years, which is the age of the MRCA of clade 3, which is older than clade 1). This revealed that Thai isolates appeared to have diverged more recently from each other than from the global isolates (Fig. 4). The majority of recent divergences in the Thai collection occurred between isolates from the same origin of isolation, i.e., environment or hospital (17 out of 24). However, there were several cases of recent divergence involving isolates of both environmental and hospital origin, as well as evidence for recent divergence between isolates from different environmental origins, i.e., different canals (Fig. 4). Moreover, one putative transmission event may have occurred between isolates of canal and farm origin (Fig. 4). The Thai isolates involved in 25 recent divergence events between a Thai and global isolates were all of clinical origin, and 15 and 7 of these isolates were recovered from invasive and non-invasive infections, respectively (Fig. 3). Furthermore, 19 out of 25 cases occurred between ST23 isolates, a well-known hyper-virulent strain. Taken together, these findings indicate that *K. pneumoniae* has rapidly expanded within the environment and hospitals,



and that in some instances, isolates of different origins have only recently diverged. Furthermore, Thai isolates were closely related to the clinical sub-population of the global collection, suggesting that the clinical collection is a part of a global circulation of hyper-virulent *K. pneumoniae* (Fig. 3).

We then investigated the distribution of plasmid replicons and virulence factors in the Thai isolates. The predominant plasmid replicons were KpN3 and ColMG828 and less frequently the R plasmid, all of which are known to carry multiple resistance and virulence genes (Additional file 2: Figure S1). The majority of the global *K. pneumoniae* isolates also harbored these plasmid replicons (results not shown), indicating their global distribution. These replicons were present in isolates from both the environment and clinical samples (Additional file 2: Figure S1). Of the 75 virulence factors considered, 10 were present in >95% of isolates, examples being the *mrk* genes encoding fimbrial biosynthesis proteins and *iutA* encoding ferric aerobactin receptor [32]. By contrast, more than 40% of virulence genes were exclusively

present in the clinical isolate collection (Additional file 2: Figure S2), including genes involved in capsule synthesis (*rmpA*) [33], iron transport (*iro*), phospholipid transport (*mce*), regulation and transport (multiple *clb* genes), and transcription regulation (*kvgA*). Multiple iron metabolism-related and siderophore genes including *ybt*, *irp*, and *fyuA* were more common (statistical significance level for z-statistic of logistic regression: *p* value < 0.05) in clinical isolates and were incorporated into the chromosome. The higher number of virulence factors in clinical versus environmental isolates (Additional file 2: Figure S2), especially those involved in iron uptake, is suggestive of hyper-mucoviscous and hyper-virulent strains that may be more efficient in iron uptake and capsule production [34]. Some virulence genes were integrated into the chromosome while others were plasmid-mediated, implying that multiple mechanisms mediate their acquisition. Some virulence genes were also present in the genome of other gram-negative bacteria such as *E. coli* and *Citrobacter koseri*, indicating sharing both within and between species. A full list of the virulence factors



with their frequencies in clinical and environmental isolates is provided in Additional file 3.

With the exception of carbapenems, trimethoprim, tige-cycline, and aminoglycosides, the Thai isolates exhibited intermediate to high resistance levels to the antibiotics tested. This finding, along with a correlation between increased minimum inhibitory concentration (MIC) values and antibiotics with a similar mechanism of action, is indicative of a limited number of effective antibiotics to treat infections (Additional file 2: Figure S3 and Figure S4). All isolates were ESBL-positive based on phenotypic testing. A genome-wide screen for known ESBLs as defined by the Comprehensive Antibiotic Resistance Database (<https://card.mcmaster.ca/>) demonstrated that *bla*_{CTX-M-15}, *bla*_{SHV}, *bla*_{VEB}, and *bla*_{GES} accounted for the ESBL phenotypes in the population. *bla*_{GES} and *bla*_{VEB} gene copies were found exclusively in the environmental hospital sewer isolates and clinical isolates, respectively. By contrast, *bla*_{CTX-M-15}

and *bla*_{SHV} had been sporadically gained by both environmental and clinical isolates. Nineteen isolates (25%) were found to be multidrug-resistant (resistant to three or more drug classes), all of which were clinical isolates. Four isolates were resistant to the carbapenem drugs. Resistance in these isolates was associated with the presence of plasmid-associated genes encoding carbapenamases, specifically New Delhi metallo-β-lactamase (NDM) (in two multidrug-resistant clinical isolates) and GES (in two non-multidrug-resistant isolates from the hospital sewer). These genes appear to have been acquired in different lineages across the phylogenetic tree (Additional file 2: Figures S5 and S6). Isolation of *Klebsiella oxytoca* and *Enterobacter cloacae* from hospital sewers that harbor GES has been reported previously [35] and indirectly reflects its presence in the hospital population or environment. Additional file 4 provides a list of β-lactamases and ESBLs in the collection.

Our results indicated that isolates with an increase in MIC values and resistant phenotypes occurred in different lineages throughout the phylogenetic tree (Additional file 2: Figure S4). However, the MIC values for β -lactams (especially cephalosporins and aztreonam) were significantly higher in clinical isolates compared with environment isolates (Additional file 2: Figure S5). Consistent with this, clinical isolates carried more copies of β -lactamase genes (Additional file 2: Figure S6). Besides the chromosomally encoded *bla*_{LEN}, *bla*_{SHV}, and *bla*_{OKP} that were present in every isolate, *bla*_{OXA}, *bla*_{TEM}, and *bla*_{CTX} occurred throughout the tree, and these genes, in combination with other β -lactamases, were more common in clinical isolates (Additional file 2: Figure S6), which is consistent with reports from numerous countries.

No patterns were observed in the presence/absence of resistance genes for the non- β -lactam antibiotics in environmental versus hospital isolates (Additional file 2: Figure S6). The *ogx* efflux pump gene was present in every isolate and was not correlated with ciprofloxacin resistance. However, non-synonymous SNP densities within DNA topoisomerase IV genes (*parE* and *parC*) and point mutations in D87 DNA gyrase A, as well as in the E84 DNA topoisomerase IV genes, were exclusively present in two isolates with high MIC values for ciprofloxacin. Both mutations occurred in the quinolone resistance-determining region (QRDR) of *gyrA* and *parC* [36, 37]. These point mutations were only found in clinical isolates, although both have been detected previously in *E. coli* isolated from aquatic environments [38]. The aminoglycoside resistance genes were noted to have been acquired on multiple occasions throughout the phylogenetic tree. Even though *K. pneumoniae* is reported to be intrinsically resistant to tetracyclines, the acquisition of further copies of tetracycline resistance genes occurred across the tree in both clinical and environmental isolates (Additional file 2: Figure S6).

Discussion

In this study we investigated the epidemiology of *K. pneumoniae* in a defined geographic area that included a general hospital and surrounding canals and farms. Our findings support the suggestion that clinical *K. pneumoniae* have evolved mechanisms to better adapt to survival in the clinical setting. Virulence factors were more frequent in clinical isolates and had been acquired on more than one occasion. This indicates the selection and dissemination of virulent strains in hospitals. Furthermore, the higher number of β -lactam resistance genes in clinical isolates together with higher absolute MIC values for some β -lactams for *K. pneumoniae* of clinical origin can be attributed to higher exposure to antibiotics, as proposed previously [39]. Given our finding of

recent divergence events between clinical and environmental isolates, our results suggest that the selective pressure imposed upon clinical isolates is sufficient to result in significant changes in the genome of clinical isolates within a few decades. It has been reported previously that the prevalence of antibiotic resistance of food-borne isolates is higher for meat-source *K. pneumoniae* isolates than for human clinical isolates [16]. However, due to the lack of sufficient data about the strength of selective antibiotic pressure in food-animal production versus hospitals, it is not possible to draw a definitive conclusion about antibiotic use and resistance [16].

The detection of carbapenem-resistant isolates in pre-treated hospital wastewater reiterates the importance of the treatment of hospital wastewater prior to release into the environment [40, 41]. Our study also highlights the role of environmental water as a potential reservoir for *K. pneumoniae*, where antibiotic resistant isolates may emerge. Antibiotic resistance genes present in environmental isolates were present in the founder lineage in some cases, for instance, in isolates in the mixed environmental and clinical clade 3 (Fig. 3 and Additional file 2: Figure S3), which is consistent with contamination of such reservoirs. Resistance was also noted to have emerged in some environmental lineages against various antibiotics, for instance, ertapenem, cefoxitin, amikacin, and amoxicillin-clavulanic acid in sewage water isolates (Additional file 2: Figure S3), which presumably arises in response to antibiotics in environmental wastewater, as reported previously [42, 43]. Several studies have shown the presence of highly antibiotic-resistant bacteria and resistance genes in sewage released into aquatic environments [42, 44–46]. In line with these findings, our results suggest that the release of untreated hospital sewage may play a role in the environmental emergence and spread of multiresistant pathogenic bacteria, and that wastewater (including hospital waste) warrants treatment to eliminate these organisms prior to release. This may be of particular importance in low-income rural areas and countries, where people are in greater contact with wastewater and may consume contaminated food and water containing high levels of antibiotic-resistant bacteria. Of note, wastewater is treated prior to release from Bhuddhasothorn Hospital, and no ESBL-positive *K. pneumoniae* were isolated from post-treated waste.

Conclusions

In this study the availability of epidemiologic information and the high resolution of whole genome sequencing allowed us to discover epidemiologic links between clinical and non-clinical *K. pneumoniae*. Limitations of the collection were the sparseness of the environmental

collection and the lack of non-ESBL-producing strains. This happened because our isolates were selected using media that were selective for ESBL production. The isolation of *K. pneumoniae* from highly contaminated samples is challenging without the use of selective culture, but our approach means that the environmental collection was biased towards clinically important isolates, i.e., ESBL-producing strains. This facilitated the identification of distinct genomic patterns relating to the distribution of antimicrobial resistance and virulence factor genes in clinical isolates, but the inclusion of non-ESBL isolates, particularly from environmental sites, is required to obtain a broader assessment of clinical and non-clinical populations. The inclusion of susceptible isolates from the environment and hospital may reduce the difference in antibiotic resistance observed here between clinical and environmental isolates, although the effect on virulence gene pattern is difficult to predict. Future studies based on larger and deeper collections will be required to gain a detailed understanding of global transmission of *K. pneumoniae* at different geographical scales.

Additional files

Additional file 1: Table S1. Supplemental Table S1, which includes accession numbers and the metadata for the studies on isolates here. (CSV 33 kb)

Additional file 2: Figures S1–S6. Figures and descriptions of Supplemental Figures S1–S6 mentioned throughout the manuscript. (DOCX 1927 kb)

Additional file 3: Table S2. Supplemental Table S2, which includes the list of putative virulence factor genes. (CSV 9 kb)

Additional file 4: Table S3. Supplemental Table S3, which includes the list of β -lactamase genes and ESBLs. (CSV 5 kb)

Abbreviations

CI: Confidence interval; CLSI: Clinical and Laboratory Standards Institute; ENA: European Nucleotide Archive; ESBL: Extended-spectrum β -lactamase; *K. pneumoniae*: *Klebsiella pneumoniae*; MIC: Minimum inhibitory concentration; MLST: multilocus sequence typing; MS: Mass spectrometry

Acknowledgements

We thank Plamena Naydenova for laboratory assistance, and the library construction, sequencing, and core informatics teams at the Wellcome Trust Sanger Institute. We would also like to thank Plamena Naydenova and Thailand One Health University Network for laboratory assistance.

Funding

CR and NC were supported by grants from the Royal Thai Golden Jubilee studentship scheme (PHD/0116/2556), Thailand, and the Newton Fund, UK and Thailand One Health University Network. This publication also presents independent research supported by the Health Innovation Challenge Fund (HICF-T5-342 and WT098600), a parallel funding partnership between the UK Department of Health and Wellcome Trust. The views expressed in this publication are those of the authors and not necessarily those of the Department of Health or the Wellcome Trust. The funding bodies did not have any role in the design of the study or in collection, analysis, or interpretation of data.

Availability of data and materials

Sequence data have been submitted to the European Nucleotide Archive (ENA) (www.ebi.ac.uk/ena) under the accession numbers listed in Additional file 1: Table S1.

Authors' contributions

DM designed the genomic data analysis framework and developed scripts to analyze the data and to interpret the results. DM and CR undertook the genomic analysis of the data. CR, SJP, and NC designed the sampling framework, and CR collected the isolates. SJP and NC were responsible for managing the project. JP was responsible for managing the project at the Wellcome Trust Sanger Institute, Hinxton, UK. DM, SJP, and NC wrote the paper. BB, SP, JT, and SA undertook laboratory work including phenotypic susceptibility testing and DNA extraction. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Consent for publication

Since there are no details on individuals reported within the manuscript, the consent for publication is not applicable.

Ethics approval and consent to participate

The study was approved by the Ethics Committee of Buddhhasothorn Hospital (BSH-IRB007/2558) and the Faculty of Tropical Medicine, Mahidol University (MUTM 2014-086-01). The research followed the principles of the Helsinki Declaration. Written informed consent was obtained from the participants. All patient information was anonymised at source and unique ID codes were used to identify cases.

Author details

¹Department of Microbiology and Immunology, Faculty of Tropical Medicine, Mahidol University, Bangkok 10400, Thailand. ²Wellcome Trust Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridgeshire, UK. ³Department of Medicine, University of Cambridge, Addenbrooke's Hospital, Box 157 Hills Road, Cambridge CB2 0QQ, UK. ⁴Division of Clinical Microbiology, Medical Technology Department, Buddhhasothorn hospital, Chachoengsao 24000, Thailand. ⁵Mahidol-Oxford Tropical Medicine Research Unit, Faculty of Tropical Medicine, Mahidol University, Bangkok 10400, Thailand. ⁶London School of Hygiene and Tropical Medicine, London WC1E 7HT, UK.

Received: 11 August 2016 Accepted: 3 January 2017

Published online: 24 January 2017

References

- Hart CA. Klebsiellae and neonates. *J Hosp Infect.* 1993;23(2):83–6.
- Podschun R, Ullmann U. *Klebsiella* spp. as nosocomial pathogens: epidemiology, taxonomy, typing methods, and pathogenicity factors. *Clin Microbiol Rev.* 1998;11(4):589–603.
- Montgomerie JZ. Epidemiology of *Klebsiella* and hospital-associated infections. *Rev Infect Dis.* 1979;1(5):736–53.
- Pitout JD, Thomson KS, Hanson ND, Ehrhardt AF, Moland ES, Sanders CC. beta-Lactamases responsible for resistance to expanded-spectrum cephalosporins in *Klebsiella pneumoniae*, *Escherichia coli*, and *Proteus mirabilis* isolates recovered in South Africa. *Antimicrob Agents Chemother.* 1998;42(6):1350–4.
- Mills JP, Talati NJ, Alby K, Han JH. The epidemiology of carbapenem-resistant *Klebsiella pneumoniae* colonization and infection among long-term acute care hospital residents. *Infect Cont Hosp Ep.* 2016;37(1):55–60.
- Coovadia YM, Johnson AP, Bhana RH, Hutchinson GR, George RC, Hafferjee IE. Multiresistant *Klebsiella-Pneumoniae* in a neonatal nursery — the importance of maintenance of infection control policies and procedures in the prevention of outbreaks. *J Hosp Infect.* 1992;22(3):197–205.
- Bauernfeind A, Rosenthal E, Eberlein E, Holley M, Schweighart S. Spread of *Klebsiella pneumoniae* producing Shv-5 beta-lactamase among hospitalized patients. *Infection.* 1993;21(1):18–22.
- Liu YY, Wang Y, Walsh TR, Yi LX, Zhang R, Spencer J, Doi Y, Tian GB, Dong BL, Huang XH, et al. Emergence of plasmid-mediated colistin resistance mechanism MCR-1 in animals and human beings in China: a microbiological and molecular biological study. *Lancet Infect Dis.* 2016;16(2):161–8.

9. Kusum M, Wongwanich S, Dhiraputra C, Pongpech P, Naenna P. Occurrence of extended-spectrum beta-lactamase in clinical isolates of *Klebsiella pneumoniae* in a University Hospital, Thailand. *J Med Assoc Thai*. 2004;87(9):1029–33.
10. Hawser SP, Bouchillon SK, Hoban DJ, Badal RE, Hsueh PR, Paterson DL. Emergence of high levels of extended-spectrum-beta-lactamase-producing gram-negative bacilli in the Asia-Pacific region: data from the Study for Monitoring Antimicrobial Resistance Trends (SMART) program, 2007. *Antimicrob Agents Chemother*. 2009;53(8):3280–4.
11. Bagley ST. Habitat association of *Klebsiella* species. *Infect Control*. 1985;6(2):52–8.
12. Podschun R, Ullmann U. Bacteriocin typing of environmental *Klebsiella* isolates. *Zentralbl Hyg Umweltmed*. 1993;195(1):22–6.
13. Seidler RJ, Knittel MD, Brown C. Potential pathogens in the environment: cultural reactions and nucleic acid studies on *Klebsiella pneumoniae* from clinical and environmental sources. *Appl Microbiol*. 1975;29(6):819–25.
14. Matsen JM, Spindler JA, Blosser RO. Characterization of *Klebsiella* isolates from natural receiving waters and comparison with human isolates. *Appl Microbiol*. 1974;28(4):672–8.
15. Knittel MD, Seidler RJ, Eby C, Cabe LM. Colonization of the botanical environment by *Klebsiella* isolates of pathogenic origin. *Appl Environ Microbiol*. 1977;34(5):557–63.
16. Davis GS, Waits K, Nordstrom L, Weaver B, Aziz M, Gauld L, Grande H, Bigler R, Horwinski J, Porter S, et al. Intermingled *Klebsiella pneumoniae* populations between retail meats and human urinary tract infections. *Clin Infect Dis*. 2015;61(6):892–9.
17. Kahn LH. The need for one health degree programs. *Infect Ecol Epidemiol*. 2011;1.
18. Karesh WB, Dobson A, Lloyd-Smith JO, Lubroth J, Dixon MA, Bennett M, Aldrich S, Harrington T, Formenty P, Loh EH, et al. Ecology of zoonoses: natural and unnatural histories. *Lancet*. 2012;380(9857):1936–45.
19. Holt KE, Wertheim H, Zadoks RN, Baker S, Whitehouse CA, Dance D, Jenney A, Connor TR, Hsu LY, Severin J, 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(27):E3574–3581.
20. Moradigaravand D, Boinett CJ, Martin V, Peacock SJ, Parkhill J. Recent independent emergence of multiple multidrug-resistant *Serratia marcescens* clones within the United Kingdom and Ireland. *Genome Res*. 2016.
21. Zerbino DR, Birney E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res*. 2008;18(5):821–9.
22. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics*. 2014;30(14):2068–9.
23. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. Genome Project Data Processing S. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009;25(16):2078–9.
24. Harris SR, Feil EJ, Holden MT, Quail MA, Nickerson EK, Chantratita N, Gardete S, Tavares A, Day N, Lindsay JA, et al. Evolution of MRSA during hospital transmission and intercontinental spread. *Science*. 2010; 327(5964):469–74.
25. Price MN, Dehal PS, Arkin AP. FastTree 2—approximately maximum-likelihood trees for large alignments. *Plos One*. 2010;5(3):e9490.
26. Croucher NJ, Page AJ, Connor TR, Delaney AJ, Keane JA, Bentley SD, Parkhill J, Harris SR. Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. *Nucleic Acids Res*. 2015;43(3):e15.
27. Drummond AJ, Rambaut A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol*. 2007;7:214.
28. Inouye M, Dashnow H, Raven LA, Schultz MB, Pope BJ, Tomita T, Zobel J, Holt KE. SRST2: Rapid genomic surveillance for public health and hospital microbiology labs. *Genome Med*. 2014;6(11):90.
29. Marcade G, Brisse S, Bialek S, Marcon E, Leflon-Guibout V, Passet V, Moreau R, Nicolas-Chanoine MH. The emergence of multidrug-resistant *Klebsiella pneumoniae* of international clones ST13, ST16, ST35, ST48 and ST101 in a teaching hospital in the Paris region. *Epidemiol Infect*. 2013;141(8):1705–12.
30. Castanheira M, Farrell SE, Wanger A, Rolston KV, Jones RN, Mendes RE. Rapid expansion of KPC-2-producing *Klebsiella pneumoniae* isolates in two Texas hospitals due to clonal spread of ST258 and ST307 lineages. *Microb Drug Resist*. 2013;19(4):295–7.
31. Bathoorn E, Rossen JW, Lokate M, Friedrich AW, Hammerum AM. Isolation of an NDM-5-producing ST16 *Klebsiella pneumoniae* from a Dutch patient without travel history abroad, August 2015. *Eurosurveillance*. 2015;20(41):2–4.
32. Landgraf TN, Berlese A, Fernandes FF, Milanezi ML, Martinez R, Panunto-Castelo A. The ferric aerobactin receptor *lutA*, a protein isolated on agarose column, is not essential for uropathogenic *Escherichia coli* infection. *Rev Lat Am Enfermagem*. 2012;20(2):340–5.
33. Hsu CR, Lin TL, Chen YC, Chou HC, Wang JT. The role of *Klebsiella pneumoniae* *mpA* in capsular polysaccharide synthesis and virulence revisited. *Microbiology*. 2011;157(Pt 12):3446–57.
34. Shon AS, Bajwa RP, Russo TA. Hypervirulent (hypermutoviscous) *Klebsiella pneumoniae*: a new and dangerous breed. *Virulence*. 2013;4(2):107–18.
35. White L, Hopkins KL, Meunier D, Perry CL, Pike R, Wilkinson P, Pickup RW, Cheesbrough J, Woodford N. Carbapenemase-producing Enterobacteriaceae in hospital wastewater: a reservoir that may be unrelated to clinical isolates. *J Hosp Infect*. 2016;93(2):145–51.
36. Yamagishi JI, Kojima T, Oyamada Y, Fujimoto K, Hattori H, Nakamura S, Inoue M. Alterations in the DNA topoisomerase IV *grIA* gene responsible for quinolone resistance in *Staphylococcus aureus*. *Antimicrob Agents Ch*. 1996;40(5):1157–63.
37. Zhou YY, Yu L, Li J, Zhang LJ, Tong Y, Kan B. Accumulation of mutations in DNA gyrase and topoisomerase IV genes contributes to fluoroquinolone resistance in *Vibrio cholerae* O139 strains. *Int J Antimicrob Ag*. 2013;42(1):72–5.
38. Johnning A, Kristiansson E, Fick J, Weijdegard B, Larsson DG. Resistance mutations in *gyrA* and *parC* are common in *Escherichia* communities of both fluoroquinolone-polluted and uncontaminated aquatic environments. *Front Microbiol*. 2015;6:1355.
39. Podschun R. Phenotypic properties of *Klebsiella pneumoniae* and *K. oxytoca* isolated from different sources. *Zentralbl Hyg Umweltmed*. 1990;189(6):527–35.
40. Munoz-Price LS, Poirel L, Bonomo RA, Schwaber MJ, Daikos GL, Cormican M, Cornaglia G, Garau J, Gniadkowski M, Hayden MK, et al. Clinical epidemiology of the global expansion of *Klebsiella pneumoniae* carbapenemases. *Lancet Infect Dis*. 2013;13(9):785–96.
41. Won SY, Munoz-Price LS, Lolans K, Hota B, Weinstein RA, Hayden MK, Centers for Disease C, Prevention Epicenter P. Emergence and rapid regional spread of *Klebsiella pneumoniae* carbapenemase-producing Enterobacteriaceae. *Clin Infect Dis*. 2011;53(6):532–40.
42. Rizzo L, Manaia C, Merlin C, Schwartz T, Dagot C, Ploy MC, Michael I, Fatta-Kassinos D. Urban wastewater treatment plants as hotspots for antibiotic resistant bacteria and genes spread into the environment: a review. *Sci Total Environ*. 2013;447:345–60.
43. Michael I, Rizzo L, McArdell CS, Manaia CM, Merlin C, Schwartz T, Dagot C, Fatta-Kassinos D. Urban wastewater treatment plants as hotspots for the release of antibiotics in the environment: a review. *Water Res*. 2013;47(3):957–95.
44. Kummerer K. Resistance in the environment. *J Antimicrob Chemother*. 2004; 54(2):311–20.
45. Czekalski N, Diez EG, Burgmann H. Wastewater as a point source of antibiotic-resistance genes in the sediment of a freshwater lake. *Isme J*. 2014;8(7):1381–90.
46. Santoro DO, Cardoso AM, Coutinho FH, Pinto LH, Vieira RP, Albano RM, Clementino MM. Diversity and antibiotic resistance profiles of Pseudomonads from a hospital wastewater treatment plant. *J Appl Microbiol*. 2015;119(6):1527–40.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit

