Population structure and antimicrobial resistance among *Klebsiella* isolates sampled from human, animal, and environmental sources in Ghana: a cross-sectional genomic One Health study

Jessica K Calland, Kaisa Haukka, Stephen W Kpordze, Atanyiwoen Brusah, Marta Corbella, Cristina Merla, Ørjan Samuelsen, Edward J Feil, Davide Sassera, Akosua B Karikari, Courage K S Saba*, Harry A Thorpe*, Jukka Corander**

Summary

Background One Health approaches to address the increasing threat of antimicrobial resistance (AMR) are gaining attention. However, data on the distribution and movement of bacteria and their AMR-associated genes between clinical and non-clinical sources are scarce, especially from low-income and middle-income countries. We aimed to analyse *Klebsiella* isolates from various sources in Ghana and compare the prevalence of AMR with datasets from two other countries.

Methods We conducted a cross-sectional genomic One Health study. Multiple clinical, environmental, and animal sources were sampled from 78 locations (eg, hospitals, residential areas, and farms) in and around Tamale, Ghana. Clinical samples were collected through routine screening and in cases of suspected infection between March 15 and Sept 15, 2019, and samples from the wider environment were collected during a dedicated sampling effort between the dates of Aug 19, 2018, and Sept 26, 2019. Sampling locations were approximately evenly distributed from the centre of the city and steadily outwards to capture both rural and urban locations. Samples with positive growth for *Klebsiella* were included. Isolates of *Klebsiella* were obtained from the samples using Simmons citrate agar medium and characterised by antimicrobial susceptibility testing and whole-genome sequencing. A comparative analysis with *Klebsiella* population surveys from Pavia, Italy, and Tromsø, Norway, was performed. AMR-associated and virulence genes were detected, and the population distribution of these genes was studied.

Findings Of 957 samples collected around Tamale, Ghana, 620 were positive for *Klebsiella* spp. 573 *Klebsiella* isolates were successfully sequenced, of which 370 were *Klebsiella pneumoniae*. Only two hospital isolates were carbapenem-resistant. Extended-spectrum β-lactamase (ESBL) genes were relatively common among the Ghanaian clinical isolates but rare in the environmental samples. Prevalence of ESBL genes in human—hospital disease samples was 64% (14 of 22 isolates) in Ghana and 44% (four of nine isolates) in Italy, and prevalence in human—hospital carriage samples was 7% (eight of 107) in Ghana and 13% (54 of 428) in Italy; the prevalence was higher in human—hospital disease samples than in human—hospital carriage samples in both countries, and prevalence across both samples in both countries was higher than in Norway. Ghanaian isolates showed evidence of high recombination rates (recombination events compared with point mutations [r/m] 9·455) and a considerable accessory gene overlap with isolates from Italy and Norway.

Interpretation Although several AMR-associated gene classes were observed relatively frequently in non-clinical sources, ESBL, carbapenemase, and virulence genes were predominantly present only in hospital samples. These results suggest that interventions should be focused on clinical settings to have the greatest effect on the prevalence and dissemination of AMR-associated genes.

Funding European Research Council (742158), Academy of Finland EuroHPC grant, Trond Mohn Foundation (BATTALION grant), and Wellcome Trust.

Introduction

Klebsiella pneumoniae is a major human pathogen with the ability to cause a wide range of nosocomial infections, including pneumonia, wound infections, soft tissue infections, and urinary tract infections, and is a leading cause of neonatal sepsis. The pathogen also causes serious community-associated infections, including pyogenic liver abscesses, pneumonia, and meningitis.[1,2] The steady increase in antimicrobial resistance (AMR) has resulted in extended-spectrum β-lactamase (ESBL)-producing and carbapenemase-producing *K pneumoniae* being included in the WHO global priority pathogen list.[3] The increase in prevalence of carbapenemase-producing *K pneumoniae* is of

Lancet Microbe 2023; 4: e943–52

Published Online October 16, 2023 https://doi.org/10.1016/S2666-5247(23)00208-2

*Contributed equally

Oslo Centre for Biostatistics and Epidemiology, Oslo University Hospital, Oslo, Norway (J K Calland PhD); **Department of Microbiology** (K Haukka PhD) **and Department of Mathematics and Statistics** (Prof J Corander PhD), **University of Helsinki, Helsinki, Finland; Department of Biotechnology** (S W Kpordze MPhil, A Brusah BSc) **and Department of Clinical Microbiology** (A B Karikari PhD), **University for Development Studies, Tamale, Ghana; Struttura Operativa Complessa di Microbiologia e Virologia, Fondazione IRCCS Policlinico San Matteo, Pavia, Italy** (M Corbella PhD, C Merla PhD); **Norwegian National Advisory Unit on Detection of Antimicrobial Resistance, Department of Microbiology and Infection Control, University Hospital of North Norway, Tromsø, Norway** (Prof Ø Samuelsen PhD); **Department of Pharmacy, UiT The Arctic University of Norway, Tromsø, Norway** (Prof Ø Samuelsen); **Milner Centre for Evolution, University of Bath, Bath, UK** (Prof E J Feil PhD); **Department of Biology and Biotechnology, University of Pavia, Pavia, Italy** (Prof D Sassera PhD); **Department of Microbiology, Faculty of Biosciences, University for Development Studies, Tamale, Ghana** (C K S Saba PhD); **Oslo Centre for Biostatistics and Epidemiology, University of Oslo, Oslo, Norway** (H A Thorpe PhD, Prof J Corander); **Parasites and Microbes, Wellcome Sanger Institute, Cambridge, UK** (Prof J Corander)

Correspondence to:
Dr Jessica K Calland, Oslo Centre for Biostatistics and Epidemiology, Oslo University Hospital, Oslo 0317, Norway
jessica.calland@medisin.uio.no

Research in context

Evidence before this study

We searched PubMed from database inception to Jan 15, 2023, without language restrictions using the terms "*Klebsiella pneumoniae*" AND "antimicrobial resistance" AND "Ghana" AND "whole genome sequencing". This search returned four publications describing the genomic antimicrobial resistance (AMR) profiles of *Klebsiella* isolates found in clinical settings. Of these studies, three were done in Ghana and the remaining study was in India. All Ghanaian studies identified the high prevalence of either carbapenemase or extended-spectrum β-lactamase (ESBL) genes. However, sampling was not extended beyond clinical settings to other potential ecological niches. AMR in bacteria from the genus *Klebsiella* is recognised as a threat to public health globally, and surveillance of these species has become a key research priority in many high-income countries. In particular, carbapenemase-producing and ESBL-producing *Klebsiella pneumoniae* are recognised by WHO as critical priority antimicrobial-resistant pathogens. Intensive genomic surveillance of *Klebsiella* from hospitals, animals, and the environment has indicated nosocomial transmission as the primary contributor to the wide dissemination of multidrug-resistant *Klebsiella* spp in high-income countries. However, evidence for the relative importance of nosocomial factors versus other ecological drivers in low-income and-middle-income countries (LMICs) is insufficient due to absence of genomic surveillance studies. Hence, this gap in knowledge needs to be urgently filled to inform future AMR policy and research priorities.

Added value of this study

Our One Health genomic survey of *Klebsiella* in Ghana allowed a detailed assessment of both the prevalence of multidrug-resistant *Klebsiella* in an LMIC setting and the relative importance of different ecological factors in their success. Through a comparative analysis of the Ghanaian isolates against earlier genomic population surveys conducted in Pavia, Italy, and Tromsø, Norway, we could set the LMIC observations in an international context that allowed for improved generalised conclusions.

Implications of all the available evidence

Our findings emphasise the importance of the clinical use of antibiotics, such as carbapenems, as a key driver of success for carbapenem-resistant *Klebsiella pneumoniae*, even in an LMIC setting. Combined evidence from the frequent horizontal gene transfer observed in *K pneumoniae* lineages in Ghana and the relatively low prevalence of ESBL genes outside the clinical setting suggests that isolates carrying these AMR-associated genes are less successful outside the hospital than in the hospital setting. These results suggest that interventions should be focused on clinical settings to have most impact on the prevalence and dissemination of AMR-associated genes.

particular concern, because carbapenems are among the last-line options for treating infections.

The best way to deal with this increase in AMR is the subject of much debate, with one strategy being to use a One Health approach by working with agricultural, veterinary, and environmental sectors in addition to human sectors to try to reduce the public health burden of antimicrobial-resistant infections.[4] The relevance of this approach depends on the interconnectedness of these different environments and the degree to which bacteria are able to move between, adapt to, and thrive in them, factors that can vary between different bacterial species and genera. The *Klebsiella* genus is genetically and ecologically diverse, consisting of 17 recognised species,[2] which are divided into four species groups, each named after the canonical species of the group.[5] *Klebsiella pneumoniae* (sensu stricto) is the most clinically important species, but most species can be isolated from many different environments, and these characteristics suggest that the genus is a good candidate for AMR studies using a One Health approach.[1,2,5] The *K pneumoniae* population is structured, with most AMR restricted to a few hospital-associated *K pneumoniae* clones.[6] The severity of *K pneumoniae* infections is partly determined by virulence factors that are generally carried on plasmids, and lineages that cause severe infections tend to be community associated.[2] The virulent and resistant clones are largely non-overlapping, although some convergent antimicrobial-resistant–virulent isolates have been observed.[7–9]

A few studies have attempted to assess the contribution of different sectors to the public health burden of antimicrobial-resistant *K pneumoniae* by sampling *Klebsiella* from many different clinical and non-clinical sources within a small temporal and geographical range and then using genomics to identify the presence of AMR and the frequency of transmission between different sectors.[5,10–12] Crucially, these studies used culture conditions that were not selective for AMR to reduce biases in AMR prevalence estimates in different sectors. Such studies have shown that AMR is overwhelmingly concentrated in hospital settings and that transmission between environmental and clinical settings is much less common than transmission within clinical settings.

Compared with high-income countries, few data have been collected from low-income and middle-income countries (LMICs), and such data are crucial to understand the spread of AMR globally.[13] Several conditions in LMICs are often cited as risk factors for the spread of AMR, including people living in close proximity to animals, often with poor hygiene and sanitation, and weak regulation of antibiotic use. By contrast, in high-income countries, the availability of antibiotics is more strongly controlled and campaigns focused on antibiotic stewardship are

implemented to restrict and regulate their usage. However, there is also greater usage of last-line intravenous antibiotics (eg, carbapenems) than in LMICs, which creates a selective environment for resistance in the clinical sector. Here, we aimed to analyse *Klebsiella* isolates from human, animal, and environmental sources from Tamale, Ghana, and compare the population structures and AMR profiles of these isolates with similar collections from Pavia, Italy,[5] and Tromsø, Norway.[14] This comparative analysis enabled us to compare the prevalence of AMR in a clinically important pathogen in these three different countries in a One Health context.

Methods

Study design and sampling

We conducted a cross-sectional genomic One Health study. We collected samples from 78 locations (eg, hospitals, residential areas, and farms; appendix 1 p 1) in Tamale, Ghana. The sampling locations were chosen to capture as many rural and urban points of potential Klebsiella transmission within and around Tamale as possible. Sample locations were approximately evenly distributed from the centre of the city and steadily outwards to more rural areas. Clinical samples were collected through routine screening and in cases of suspected infection between March 15 and Sept 15, 2019, and samples from the wider environment were collected during a dedicated sampling effort between Aug 19, 2018, and Sept 26, 2019. All samples with positive growth for *Klebsiella* were included. Sample sources were initially divided into nine major categories: human—community carriage (ie, faecal samples from healthy people in the community), human—hospital carriage (ie, faecal samples from routine hospital screening, with no suspected *Klebsiella* infection), human—hospital disease (ie, samples from patients with suspected infection from a wound, urine, vagina, or sputum), chickens, environmental surfaces, farm animals, invertebrates, water, and other (appendix 1 p 1). For a more detailed understanding of AMR prevalence in different niches, an additional 16 subcategories were established for the Ghananian dataset (appendix 1 pp 1–3).

This study was approved by Tamale Teaching Hospital Head of Ethics and Head of Laboratory under reference number TTH/R&D/SR/158 for a period of 13 months (ie, Aug 1, 2018, to Sept 30, 2019). Bacterial samples were collected routinely from patients and consent was therefore waived. Community carriage samples were collected from outdoor defecation sites under the same approval as the hospital patient samples. Genomic data were de-identified when reported as findings in the study. Verbal consent was sought from farm owners for the collection of bacterial samples from animals.

Procedures

Hospital human samples were collected by hospital staff according to screening and infection control protocols. All other environmental (including hospital environment), farm, and animal samples were collected by the authors (KH, SWK, AB, ABK, and CKSS) from multiple sources and locations in and around Tamale. After collection, samples were stored at –80°C and then transported to Italy on dry ice. Samples were processed according to the same protocols used by Thorpe and colleagues for the Italian study.[5] Samples were pre-enriched in Luria-Bertani broth with amoxicillin (10 mg/mL), incubated for 18–24 h at 37°C ± 1°C, and plated onto Simmons citrate agar with inositol (SCAI) medium[15,16] and incubated at 36 ± 1°C for 48 h. Yellow and mucoid colonies on SCAI plates were identified at the species level through matrix-assisted laser desorption ionisation-time of flight and subcultured on MacConkey agar for antibiotic susceptibility testing and DNA extraction. Antibiotic susceptibility was tested with the BD Phoenix 100 automated system (Becton, Dickinson and Company, Franklin Lakes, NJ, USA) with the NMIC-417 panel, and using the European Committee on Antimicrobial Susceptibility Testing breakpoints.[17] Genomic DNA was extracted using a QIAsymphony instrument (Qiagen, Milan, Italy) and the QIAsymphony DSP Virus/Pathogen kit (Qiagen, Milan, Italy). Genomic DNA libraries were prepared with the Nextera XT Library Prep Kit (Illumina, San Diego, CA, USA). Illumina sequencing was performed at the Wellcome Trust Sanger Institute with HiSeq X10, 150 base pair, paired-end reads. Reads were trimmed with Trimmomatic version 0.33,[18] and SPAdes version 3.9.0 was used to generate de novo assemblies.[19] Assemblies were annotated with Prokka version 1.12 with default parameters.[20]

See Online for appendix 1

Because *K pneumoniae* is considered the most clinically important *Klebsiella* species, the Ghanaian dataset was combined with two other geographically distinct *K pneumoniae* datasets for comparative analyses (1127 *K pneumoniae* isolates from human, animal, and environmental sources sampled in Pavia, Italy,[5] and 303 *K pneumoniae* from a screening study of healthy people in Tromsø, Norway;[14] appendix 1 p 3). Kleborate version 2.1.0 was used to group the isolates into *Klebsiella* species, assign multilocus sequence types (STs) to species from the *K pneumoniae* species complex, and detect AMR-associated and virulence genes.[8] Isolates were divided into sequence clusters (SCs) with PopPUNK version 2.0.2.[21]

A whole-genus neighbour-joining phylogeny of the isolates from Ghana was constructed by use of RapidNJ[22] from a matrix of Mash distances.[23] A separate neighbour-joining phylogeny of the three comparative *K pneumoniae* datasets was constructed from the core-genome distances estimated by PopPUNK version 2.0.2.[21] Individual core-genome alignments for the top ten most common SCs across the three datasets were constructed by mapping assemblies to the complete reference genome NTUH-K2044 (accession number NC_016845.1) using the BWA-MEM algorithm[24] and snippy-core using

Snippy version 4.6.0.[25] Individual approximate maximum-likelihood phylogenies were built with FastTree version 2.1.8 (double precision)[26] using the generalised time-reversible model of nucleotide evolution (appendix 2 p 5). Gubbins version 3.1.6 (default settings)[27] was used to infer the effects of recombination on the nine individual SC alignments (excluding SC9_ST3068, because this SC was present only in the Italian study;[5] appendix 2 p 6) and estimate the predicted ratio of nucleotide changes imported through recombination events compared with point mutation (r/m). Assemblies of *K pneumoniae* from the three independent datasets were annotated with Prokka[20] version 1.13 and the pangenome was estimated using Panaroo (strict mode) version 1.3.0.[28] Pangenome graphs were merged and clustering of accessory gene content visualised with stochastic cluster embedding implemented by Mandrake version 1.2.2.[29] Further details on study design, bacterial processing, DNA extraction, and genomic analysis are shown in appendix 2 (pp 1–4).

See Online for appendix 2

For more on **rstatix** see https://cran.r-project.org/web/packages/rstatix/index.html

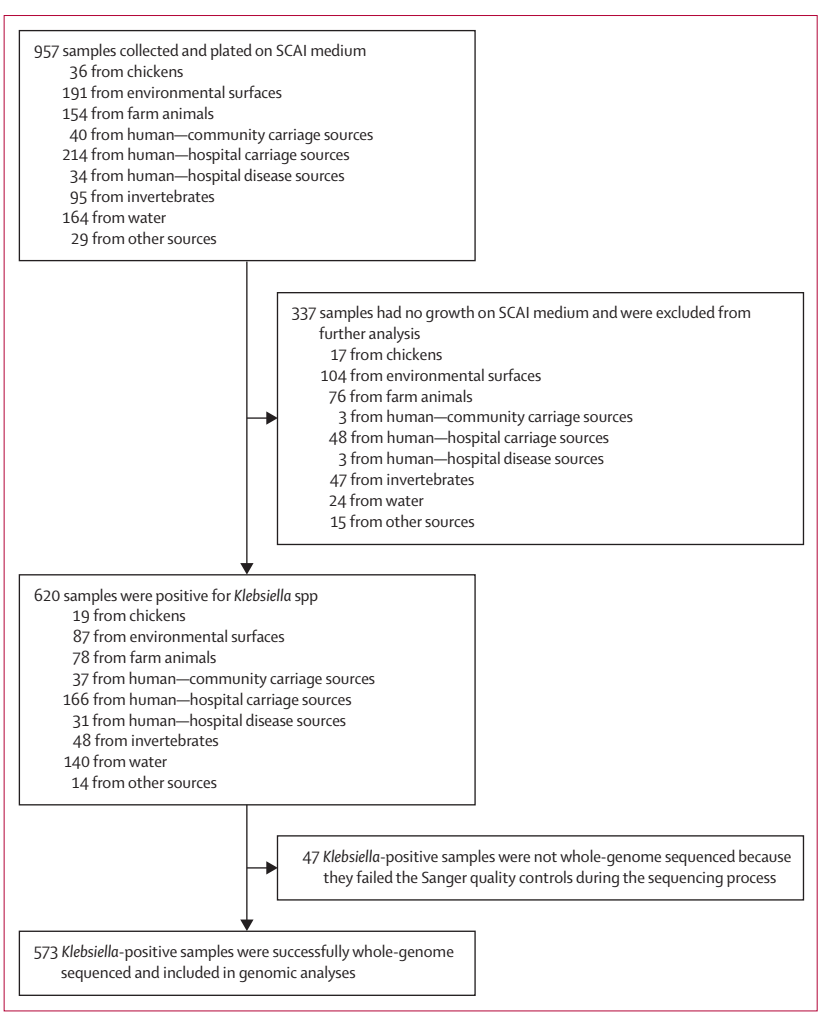

Figure 1: **Study flowchart**
SCAI=Simmons citrate agar with inositol.

Statistical analysis

The main outcomes of this study were the characterisation of the distribution of *Klebsiella* species in Italy and Ghana and across sample sources and of *K pneumoniae* SCs, STs, AMR-associated genes, and virulence genes in the Ghanian, Italian, and Norwegian datasets. An additional outcome was the recombination rate across *K pneumoniae* SCs.

The statistical significance of the distribution of all STs between the three datasets was compared by computing a Pearson's χ^2 test. A simulated p value (significance determined by $p<0{\cdot}05$) was computed for a Monte Carlo test with 10 000 replicates to account for missing data. We used rstatix package, version 0.7.2.

Role of the funding source

The funders of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report.

Results

957 samples were collected around the city of Tamale, Ghana, from 78 locations (figure 1, 2; appendix 1 p 1), between Aug 19, 2018, and Sept 26, 2019. 620 samples were positive for *Klebsiella* spp and the genomes of 573 single colonies were successfully sequenced (appendix 1 pp 1–2). The source with the highest prevalence of *Klebsiella* spp growth was human—hospital carriage (166 [27%] of 620 samples), followed by water (140 [23%]), environmental surfaces (87 [14%]), and farm animal (78 [13%]). The remaining source categories contained fewer than 50 isolates each (figure 2; appendix 1 pp 1–2).

K pneumoniae was the most common species in Ghana (370 [65%] of 573 *Klebsiella* whole-genome sequences). *Klebsiella quasipneumoniae* subspecies *similipneumoniae* was next (136 [24%]), followed by *Klebsiella variicola* subspecies *variicola* (33 [6%]), and then *Klebsiella aerogenes* (24 [4%]). Five other species were identified (ten [2%]; figure 2; appendix 1 pp 2–4).

Most of the *Klebsiella* species identified in Ghana were also found in the Italian dataset,[5] except for *Klebsiella africana* and *Klebsiella* Ka4, which is a newly identified cluster of four isolates that are closely related to *K aerogenes*.[8] The Italian dataset contained an additional nine species that were not present in the Ghanaian dataset (appendix 1 p 4). *K pneumoniae* was the most frequently sampled species in Italy (1705 [49%] of 3483 isolates), followed by *Klebsiella michiganensis* (300 [9%]), *K variicola* subspecies *variicola* (279 [8%]), and *Klebsiella ornithinolytica* (258 [7%]; appendix 1 p 4). *K quasipneumoniae* subspecies *similipneumoniae* was much more common in Ghana (136 [24%] of 573) than in Italy (49 [1%] of 3483).[5]

Little source specificity was identified in the Ghanaian isolates; *K pneumoniae* and *K quasipneumoniae*

subspecies *similipneumoniae* were isolated from all sources, whereas *K aerogenes* was isolated from eight of nine sources and *K variicola* subspecies *variicola* was isolated from seven of nine sources. However, some species were more frequently sampled from a particular source than others. For example, *K pneumoniae* was frequently observed in human—hospital carriage samples (107 [72%] of 148 isolates) and water sources (81 [58%] of 140 isolates), and *K quasipneumoniae* subspecies *similipneumoniae* isolates were also common in water sources (50 [36%] of 140 isolates; figure 2; appendix 1 pp 1–2).

The most abundant species identified was *K pneumoniae*, which is also considered the most clinically important *Klebsiella* species.[2] For these reasons, the remainder of genomic analyses focused on *K pneumoniae* identified from this study (n=370), contextualised with the data from the Italian study[5] (n=1127) and the Norwegian study[14] (n=303; figure 3; appendix 1 p 3). Two additional sources (ie, dogs and cats and plants) were added to reflect the broader sampling in the Italian study. Isolates from the three studies were split into 532 genetically distinct SCs. Differences in the distribution of STs or SCs between the three datasets were significant (Pearson's χ^2, p<0·0001). The distribution of isolates among the SCs was skewed; 910 (51%) of 1800 isolates were present in 35 SCs, each containing at least nine isolates. 17 (49%) of 35 SCs were present in all three datasets, of which the most common were SC12 (ST200), SC14 (ST15), SC16 (ST35), SC2 (ST17), SC4 (ST45), SC5 (ST661), SC6 (ST37), SC7 (ST36), and SC8 (ST29). Notable clones of interest were well studied hospital-associated clones (SC1_ST307, 35 [2%] of 1800; SC4_45_139_356_ST45, 70 [4%]) and human-associated clones (SC2_ST17, 135 [8%]; SC7_ST36, 47 [3%]). 18 of the 35 most common SCs overall were not present in all datasets, indicating differences in the SC distribution between datasets. There was a long tail of rare SCs; 441 SCs were represented by three isolates or fewer, totalling 601 of 1800 isolates, and 430 SCs were present in only one of the datasets.

The major *K pneumoniae* SCs were isolated from a wide range of sources; eight of the ten most common SCs were isolated from at least five of 11 sources (and five were isolated from at least eight sources). However, some SCs were more frequently observed in particular sources; for example, SC1_ST307, SC2_ST17, SC4_45_139_356_ST45, and SC7_ST36 were more common in human—hospital carriage sources than in other sources, and SC5_ST661, SC6_ST348, SC9_ST3068, SC10_ST2073, and SC12_ST200 were more common in farm animal sources than in other sources (figure 3; appendix 1 p 3).

To improve understanding of the distribution of AMR-associated genes within and between the three datasets, the major source categories were subdivided into 22 categories (appendix 1 p 3, appendix 2 p 7). AMR-associated genes were concentrated in clinical sources in Ghana, with the highest levels in the human—hospital disease (6·5 AMR gene classes per isolate) and hospital environment sources (4·8 classes per isolate; appendix 2 pp 8–9). AMR-associated genes were less prevalent in the other sources (0·7 classes per isolate). The two sources with the highest total prevalence of AMR-associated genes in Italy were also the clinical sources human—hospital disease (4·3 classes per isolate) and hospital environment (9·0 classes per isolate; figure 4; appendix 2 pp 7, 8, 10). AMR-associated genes were less prevalent in non-clinical settings (ie, 19 sources) than in clinical sources, with a mean number of AMR gene classes per isolate of 0·3 in the Italian dataset and 0·2 in the Norwegian human—community carriage dataset (appendix 2 pp 8, 11).

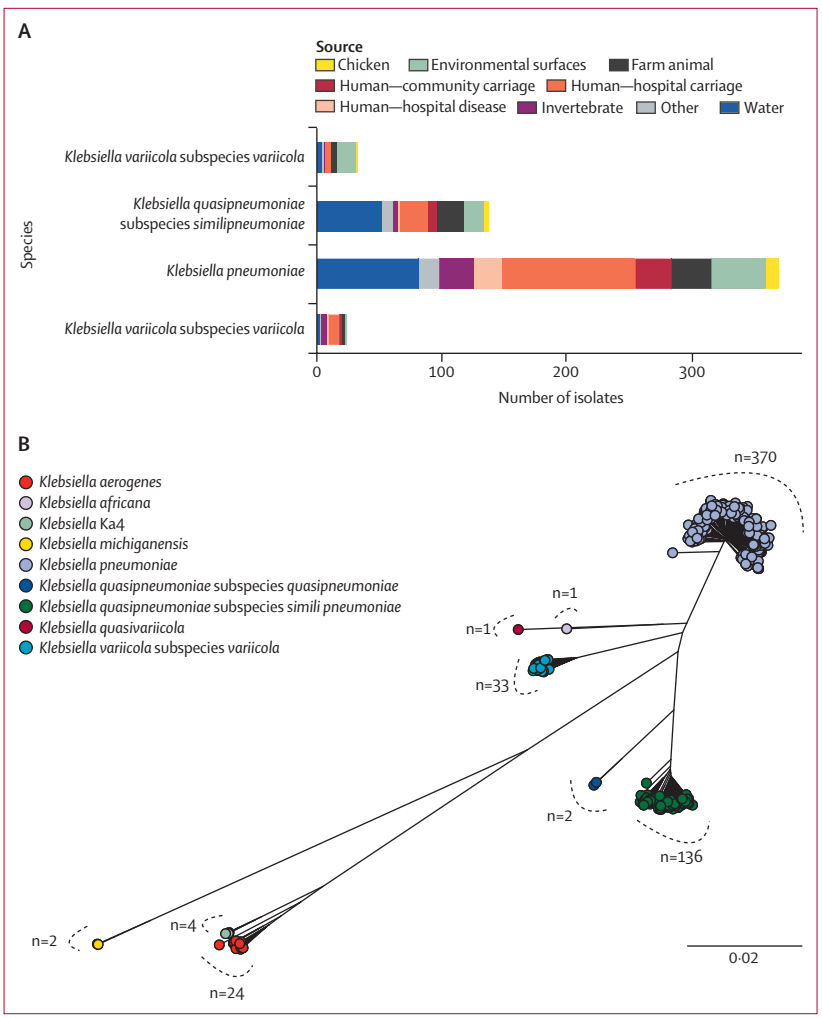

***Figure 2:* *Klebsiella* spp isolates sampled in Ghana**
(A) Bar chart for the four species containing at least ten isolates. (B) Neighbour-joining phylogenetic tree showing the population structure and isolate counts of the nine *Klebsiella* species successfully isolated and whole-genome sequenced from this study (n=573). The scale bar refers to the number of single-nucleotide polymorphisms per site.

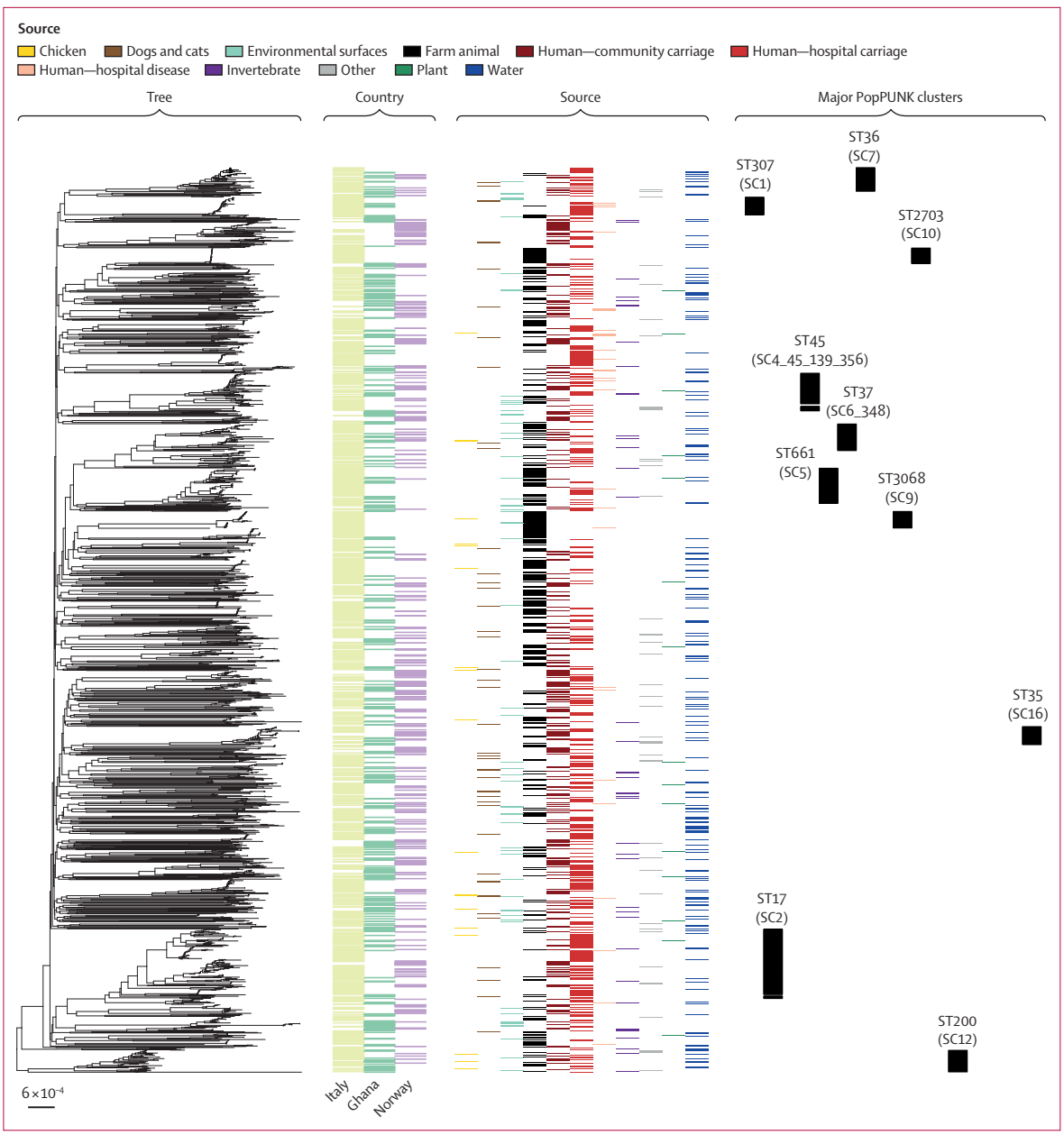

Figure 3: **Ghanaian *Klebsiella pneumoniae* in context with Italy and Norway**
Neighbour-joining phylogenetic tree of *Klebsiella pneumoniae* genomes sampled from the Ghanaian dataset (n=370) compared with *Klebsiella pneumoniae* genomes from Pavia, Italy (n=1127) and Tromsø, Norway (n=303). The scale bar refers to the number of single-nucleotide polymorphisms per site. Terminal tree nodes are given according to country and source. Sources were divided into 11 categories. Clusters of genetically distinct isolates were determined with PopPUNK.[21] The top ten most commonly shared SCs across the three datasets are shown by black tiles in the fourth panel. Each SC is shown in parentheses with the corresponding ST given above. SC=sequence cluster. ST=sequence type.

Third-generation cephalosporins and carbapenems are among the most important antibiotic classes for treating multidrug-resistant *K pneumoniae* infections, so we focused on the genes that encode enzymes that confer resistance to these antibiotics (ie, ESBLs and carbapenemases). An ESBL gene was detected in 36 isolates in Ghana, of which 29 showed phenotypic resistance to cefixime, 26 showed phenotypic resistance to ceftazidime, and 29 showed phenotypic resistance to ceftriaxone, and no phenotypic resistance was observed in the absence of a detected ESBL gene (appendix 1 p 5). These results indicated good concordance between genotype (ie, presence of resistance-associated genes) and phenotypic resistance. The most common ESBL gene was $bla_{\text{CTX-M-15}}$ (30 of 36 occurrences), with the others being $bla_{\text{CTX-M-3}}$ (two occurrences) and bla_{SHV} alleles

Figure 4: **Distribution of antimicrobial resistance and virulence within the three datasets**
Neighbour-joining phylogenetic tree of only the *Klebsiella pneumoniae* genomes sampled from the Ghanaian dataset (n=370) compared with *Klebsiella pneumoniae* genomes from Pavia, Italy (n=1127) and Tromsø, Norway (n=303). The scale bar refers to the number of SNPs per site. The presence of both antimicrobial resistance and virulence for each isolate is shown. Each column corresponds to a different antimicrobial resistance or virulence factor identified using Kleborate.[8]
ESBL=extended-spectrum β-lactamase. SNP=single-nucleotide polymorphism.

(four occurrences). The 30 isolates that carried $bla_{CTX\text{-}M\text{-}15}$ were from 17 different STs, indicating that this gene was not sequence-type specific, but 23 of 30 isolates were from hospital sources, suggesting a clinical association for this gene.

ESBL genes were present at high frequencies in *K pneumoniae* in Ghana from the clinical sources human—hospital disease (14 [64%] of 22 isolates) and hospital environment (three [60%] of five isolates). ESBL genes were absent from seven of 16 sampled sources and were identified rarely in the human—community carriage (two [7%] of 29 isolates), human—hospital carriage (eight [7%] of 107), meat processing environment (one [5%] of 19), water–domestic and agricultural (one [6%] of 18), water—human wastewater (two [3%] of 60), and sheep (one [17%] of six) sources (figures 3, 4; appendix 2 pp 7–8). ESBL genes in the Italian dataset were present in seven (41%) of 17 sources. Prevalence in human—hospital disease sources was also high (four [44%] of nine isolates), and the prevalence in human—hospital carriage sources was 13% (54 of 428 isolates; figure 4; appendix 2 pp 7–8). In the remaining sources, ESBL genes were present at similar levels to those in Ghana (one [6%] of 18 sources for cats, four [1%] of 348 sources for cows, three [10%] of 31 sources for dogs, and five [7%] of 69 sources for pigs). No ESBL genes were present in the Norwegian human—community carriage dataset.

Genes encoding carbapenemases were absent from all sources in Ghana, with the exception of two positive human—hospital disease isolates. These isolates were from different STs and carried different genes (one was ST17 and carried $bla_{OXA\text{-}181}$ and the other was ST874 and carried $bla_{OXA\text{-}48}$), and both also carried the $bla_{CTX\text{-}M\text{-}15}$ ESBL gene (appendix 2 p 12). Both isolates showed phenotypic resistance to ertapenem but were sensitive to meropenem and imipenem. No other isolates showed phenotypic resistance to a carbapenem antibiotic. Carbapenemase genes were also present only in the clinical sources human—hospital carriage (42 [10%] of 428 isolates) and human—hospital disease (one [13%] of eight isolates) in Italy. No carbapenemase genes were present in the Norwegian dataset.

Two hypervirulent isolates were identified in Ghana, one SC39_ST1364 and one SC405_ST218, both from hospital patients (one from human—hospital disease and one from human—hospital carriage sources; appendix 2 p 9). Both carried a suite of virulence gene loci, including aerobactin, salmochelin, and *rmpA2* truncations, and the SC405_ST218 isolate also carried yersiniabactin. Two other isolates (one from human—hospital carriage and one from water—human wastewater sources) carried an unknown variant of the salmochelin locus, with the human—hospital carriage isolate also carrying yersiniabactin. Yersiniabactin was the most widely distributed virulence factor, found in ten sources. Yersiniabactin was most common in the hospital environment (two [40%] of five isolates) and human—hospital disease sources (12 [55%] of 22 isolates) and was identified less frequently in other sources (≤17%).

Excluding the large number (n=42) of aerobactin-carrying porcine isolates (which were discussed separately in the Italian study and do not appear to be linked to human disease[5]), few virulent isolates (defined as Kleborate virulence score ≥2) were present in Italy (ie, three from human—hospital carriage, two from dog, two from farm environment, and one from water—farm sources) and in the human—hospital carriage isolates from Norway (n=7; appendix 2 pp 14–15). These isolates were scattered across many SCs across the three datasets (ie, two isolates from two SCs in Ghana, eight isolates from eight SCs in Italy, and seven isolates from six SCs in Norway). Yersiniabactin was present in 14 sources in Italy and, as in Ghana, prevalence was highest in the hospital environment (six [100%] of six isolates) and human—hospital disease sources (four [50%] of eight isolates). Of the sources that were sampled in both datasets, the prevalence of yersiniabactin was higher in Italy in seven sources and higher in Ghana in three sources. Prevalence in human—community carriage isolates was lower in Ghana (one [3%] of 29 isolates) than in Italy (five [23%] of 22 isolates) and Norway (39 [13%] of 303 isolates). In the Italian dataset, 11 isolates showed traits for the convergence of hypervirulence and multidrug resistance within single strains (virulence score >3 and resistance score >1). However, no evidence for convergence was detected in the Ghanaian or Norwegian datasets (appendix 1 p 3).

The mean recombination rate across all SCs was highest for the Ghanaian dataset (r/m 9·455, SD 6·317), followed by Norway (5·183, 4·869) and Italy (3·683, 2·416; appendix 1 p 6, appendix 2 p 6). Individual SC recombination rates were particularly high in Ghana compared with Italy and Norway in SCs 2 (r/m 12·414 *vs* 2·100 *vs* 4·722), 4 (12·608 *vs* 2·408 *vs* 7·237), 5 (14·115 *vs* 5·931 *vs* 0·843), and 12 (20·819 *vs* 8·182 *vs* 12·709; appendix 1 p 6). The core genome size was similar for the three datasets (3971 genes for Ghana, 4158 for Italy, and 4030 for Norway) and the pangenomes were large (24 262 genes for Ghana, 30 972 for Italy, and 22 672 for Norway; appendix 1 p 7). There was strong evidence for overlap of accessory gene content between countries, SCs, and sources, suggesting that these factors are not barriers to horizontal gene transfer and therefore the potential to share AMR-associated and virulence elements is unlikely to be limited by opportunity (appendix 2 p 16).

Discussion

The increasing level of AMR observed in clinical *Klebsiella* isolates poses a serious threat to public health, and to mitigate this threat, it is important to understand

the drivers of this resistance. As *Klebsiella* spp are ecological generalists and live in many different environments, they have the potential to spread AMR genes between different settings. A few recent One Health studies have characterised *Klebsiella* from clinical, animal, and environmental settings, and have reported that AMR is largely concentrated in clinical settings.[5,10–12] Our study, which was conducted in an LMIC setting, adds to this body of evidence by providing a large sample of contemporaneous *Klebsiella* isolates from Tamale, Ghana. We identified that AMR-associated genes were distributed unevenly across different *Klebsiella* species and sources, and there was notable variation between different classes of AMR-associated genes, where ESBL and carbapenemase genes were predominantly found in clinical settings. Pangenome analyses showed that this uneven distribution was unlikely to be due to transmission barriers because accessory genomes were shared across many sources.

Carbapenemase genes were relatively common in Italy, and were present in two isolates in Ghana, but were found only in hospital samples in both datasets (and were also not observed in human—community carriage isolates from Norway). By contrast, ESBL genes were relatively common in hospital samples in Italy and Ghana and were found sporadically outside the hospital environment. These results were also consistent with a study of European food products, in which no carbapenem-resistant and only four ESBL-resistant isolates were found among a collection of 131 *Klebsiella* isolates.[30] These observations emphasise the context-dependent nature of AMR prevalence, and it is possible that high levels of carbapenem usage are required for endemic maintenance of resistance to this class of antibiotics. Previous work has shown that there is a fitness cost associated with carrying carbapenemase-encoding plasmids.[31] Hence, it is plausible that multidrug-resistant strains are outcompeted or lose the genes once they migrate to the environmental settings, or both. For other resistance classes, the situation was more nuanced, which might reflect differences in the fitness costs associated with each class.

The dominant *K pneumoniae* SCs in Ghana varied markedly with respect to those in Italy and Norway, indicating local specialisation. Of the 35 most common SCs overall (which contained 51% of all isolates), only 17 were present in all three datasets. There was much overlap in accessory gene profiles between the datasets, and no evidence of strong geographical structure at a level that would be required to prevent the movement of AMR genes. Additionally, recombination analysis showed evidence of frequent homologous recombination, which together with the overlap in accessory gene pool suggests that opportunity for AMR-associated gene exchange is unlikely to be a limiting factor, and consequently that AMR prevalence is likely to be maintained by selection.

There were several limitations to our study. With our enrichment methodology, we did not select for antibiotic resistance, and so we might have underestimated the prevalence of AMR-associated genes. However, we chose this methodology to capture the diversity and population structure in a less biased manner than if we had used selective media. Future studies could be designed to compare methodologies for the ability to detect AMR-associated genes and the resulting population structures. A related limitation is that it was difficult to find appropriate comparator studies to use due to the small number of studies using a combination of diverse source sampling with non-antibiotic selective media. The Italian study was similar in both the sampling and enrichment approaches,[5] but the Norwegian study was similar only in terms of the enrichment approach, because only healthy human community samples were included.[14] These differences limited the possibility to perform detailed comparisons between the studies. We did not perform long-read sequencing and therefore could not reconstruct plasmids; although this did not negatively affect our ability to detect AMR-associated genes, it did limit our ability to compare the plasmids carrying these genes in the three countries. Finally, we did not have access to antibiotic usage data for the three countries, which meant we could not draw any specific conclusions about the prevalence of AMR according to differences in antibiotic usage between the three countries.

Although few *Klebsiella* One Health genomics studies have been done, their results have been largely consistent.[5,10–12] Carbapenemase genes, and to a lesser extent ESBL genes, are mostly restricted to *K pneumoniae*, with only sporadic occurrences in other *Klebsiella* species. ESBL genes are generally observed at relatively low frequencies in many environments, whereas carbapenemase genes are almost exclusively observed in clinical environments. Some evidence suggests sporadic transmission between clinical and non-clinical settings might occur,[5,10–12] but no evidence for sustained onward transmission has been identified. One clear exception is humans and their companion animals, which have been shown to carry the same hospital-associated STs.[5] Overall, the available evidence indicates that clinical settings are the hubs of both AMR and its successful transmission, even in LMICs, which provides crucial information for designing future interventions to curb the success of multidrug-resistant *Klebsiella*.

Contributors

JC designed and obtained funding for the study. EJF, HAT, JC, and ØS conceptualised the study. JKC, HAT, and JC developed the methodology. KH, SWK, AB, ABK, and CKSS collected and processed samples for the study. MC, CM, and DS performed the isolation, identification, minimum inhibitory concentration testing, and DNA extraction for all samples. JKC performed genomic analysis. JKC and HAT analysed and visualised

the data. JKC and HAT drafted the manuscript. All authors reviewed and edited the manuscript. JC and JKC accessed and verified the underlying data of the study. All authors had full access to all data in the study and accept responsibility for the decision to submit for publication.

Declaration of interests

We declare no competing interests.

Data sharing

The Ghana dataset is available to explore and download at https://microreact.org/project/b1oQG4nEdcuJc5xFvt4AEq-ghana-klebsiella-all-species and combined datasets are available to explore and download at https://microreact.org/project/trkkSb3Pg54ykvH13BkcZU-combined-klebsiella-pneumoniae. The European Nucleotide Archive accession numbers for the sequencing reads generated in this study are available in appendix 1. Raw Illumina sequence data for the Norwegian and Italian studies are publicly available under the BioProject codes PRJEB42350 and PRJEB27342, respectively.

Acknowledgments

This work was supported by the European Research Council (742158), Academy of Finland EuroHPC grant, Trond Mohn Foundation (BATTALION grant), and the Wellcome Trust.

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 USA* 2015; **112:** E3574–81.

2 Wyres KL, Lam MMC, Holt KE. Population genomics of *Klebsiella pneumoniae*. *Nat Rev Microbiol* 2020; published online Feb 13. https://doi.org/10.1038/s41579-019-0315-1.

3 Tacconelli E, Carrara E, Savoldi A, et al. Discovery, research, and development of new antibiotics: the WHO priority list of antibiotic-resistant bacteria and tuberculosis. *Lancet Infect Dis* 2018; **18:** 318–27.

4 Walsh TR. A One-Health approach to antimicrobial resistance. *Nat Microbiol* 2018; **3:** 854–55.

5 Thorpe HA, Booton R, Kallonen T, et al. A large-scale genomic snapshot of *Klebsiella* spp isolates in northern Italy reveals limited transmission between clinical and non-clinical settings. *Nat Microbiol* 2022; **7:** 2054–67.

6 David S, Reuter S, Harris SR, et al. Epidemic of carbapenem-resistant *Klebsiella pneumoniae* in Europe is driven by nosocomial spread. *Nat Microbiol* 2019; **4:** 1919–29.

7 Lam MMC, Wyres KL, Wick RR, et al. Convergence of virulence and MDR in a single plasmid vector in MDR *Klebsiella pneumoniae* ST15. *J Antimicrob Chemother* 2019; **74:** 1218–22.

8 Lam MMC, Wick RR, Watts SC, Cerdeira LT, Wyres KL, Holt KE. A genomic surveillance framework and genotyping tool for *Klebsiella pneumoniae* and its related species complex. *Nat Commun* 2021; **12:** 4188.

9 Wyres KL, Nguyen TNT, Lam MMC, et al. Genomic surveillance for hypervirulence and multi-drug resistance in invasive *Klebsiella pneumoniae* from South and Southeast Asia. *Genome Med* 2020; **12:** 11.

10 Dereeper A, Gruel G, Pot M, et al. Limited transmission of *Klebsiella pneumoniae* among humans, animals, and the environment in a Caribbean island, Guadeloupe (French West Indies). *Microbiol Spectr* 2022; **10:** e0124222.

11 Ludden C, Moradigaravand D, Jamrozy D, et al. A One Health study of the genetic relatedness of *Klebsiella pneumoniae* and their mobile elements in the east of England. *Clin Infect Dis* 2020; **70:** 219–26.

12 Butaye P, Stegger M, Moodley A, et al. One Health genomic study of human and animal *Klebsiella pneumoniae* isolated at diagnostic laboratories on a small Caribbean island. *Antibiotics (Basel)* 2021; **11:** 42.

13 Ikhimiukor OO, Odih EE, Donado-Godoy P, Okeke IN. A bottom-up view of antimicrobial resistance transmission in developing countries. *Nat Microbiol* 2022; **7:** 757–65.

14 Raffelsberger N, Hetland MAK, Svendsen K, et al. Gastrointestinal carriage of *Klebsiella pneumoniae* in a general adult population: a cross-sectional study of risk factors and bacterial genomic diversity. *Gut Microbes* 2021; **13:** 1939599.

15 Van Kregten E, Westerdaal NA, Willers JM. New, simple medium for selective recovery of *Klebsiella pneumoniae* and *Klebsiella oxytoca* from human feces. *J Clin Microbiol* 1984; **20:** 936–41.

16 Passet V, Brisse S. Association of tellurite resistance with hypervirulent clonal groups of *Klebsiella pneumoniae*. *J Clin Microbiol* 2015; **53:** 1380–82.

17 European Committee on Antimicrobial Susceptibility Testing. Breakpoint tables for interpretation of MICs and zone diameters. Jan 1, 2019. https://www.eucast.org/fileadmin/src/media/PDFs/EUCAST_files/Breakpoint_tables/v_9.0_Breakpoint_Tables.pdf (accessed June 19, 2023).

18 Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 2014; **30:** 2114–20.

19 Bankevich A, Nurk S, Antipov D, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 2012; **19:** 455–77.

20 Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 2014; **30:** 2068–69.

21 Lees JA, Harris SR, Tonkin-Hill G, et al. Fast and flexible bacterial genomic epidemiology with PopPUNK. *Genome Res* 2019; **29:** 304–16.

22 Simonsen M, Mailund T, Pedersen CNS. Rapid neighbour-joining. In: Crandall KA, Lagergren J, eds. Algorithms in bioinformatics. Berlin: Springer Berlin Heidelberg, 2008: 113–22.

23 Ondov BD, Treangen TJ, Melsted P, et al. Mash: fast genome and metagenome distance estimation using MinHash. *Genome Biol* 2016; **17:** 132.

24 Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv* 2013; published online March 16. http://arxiv.org/abs/1303.3997 (preprint).

25 Seemann T. Rapid haploid variant calling and core genome alignment. https://github.com/tseemann/snippy (accessed June 15, 2022).

26 Price MN, Dehal PS, Arkin AP. FastTree 2—approximately maximum-likelihood trees for large alignments. *PLoS One* 2010; **5:** e9490.

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

28 Tonkin-Hill G, MacAlasdair N, Ruis C, et al. Producing polished prokaryotic pangenomes with the Panaroo pipeline. *Genome Biol* 2020; **21:** 180.

29 Lees JA, Tonkin-Hill G, Yang Z, Corander J. Mandrake: visualizing microbial population structure by embedding millions of genomes into a low-dimensional representation. *Philos Trans R Soc Lond B Biol Sci* 2022; **377:** 20210237.

30 Rodrigues C, Hauser K, Cahill N, et al. High prevalence of *Klebsiella pneumoniae* in European food products: a multicentric study comparing culture and molecular detection methods. *Microbiol Spectr* 2022; **10:** e0237621.

31 Kloos J, Gama JA, Hegstad J, Samuelsen Ø, Johnsen PJ. Piggybacking on niche adaptation improves the maintenance of multidrug-resistance plasmids. *Mol Biol Evol* 2021; **38:** 3188–201.