ACCESS MICROBIOLOGY
an open research platform

RESEARCH ARTICLE
Dinda *et al.*, *Access Microbiology* 2024;6:000667.v4
DOI 10.1099/acmi.0.000667.v4

MICROBIOLOGY SOCIETY
OPEN DATA OPEN ACCESS

Whole genome sequencing and genotyping *Klebsiella pneumoniae* multi-drug resistant hospital isolates from Western Kenya

Victor Dinda[1], Andrew Nyerere Kimang'a[2], Daniel Kariuki[3], Anthony Wawire Sifuna[4], Thomas James O'Brien[5], Martin Welch[5,*] and Oleg N. Reva[6,*]

Abstract

Objectives. *Klebsiella pneumoniae* are a frequent cause of nosocomial infections worldwide. Sequence type 147 (ST147) has been reported as a major circulating high-risk lineage in many countries, and appears to be a formidable platform for the dissemination of antimicrobial resistance (AMR) determinants. However, the distribution of this pathogen in Western African hospitals has been scarcely studied. The main objective of this work was to perform whole genome sequencing of *K. pneumoniae* isolates from a referral hospital in Kakamega (Kenya) for genotyping and identification of AMR and virulence determinants.

Methods. In total, 15 *K. pneumoniae* isolates showing a broad spectrum antimicrobial resistance were selected for whole genome sequencing by Illumina HiSeq 2500 platform.

Results. ST147 was the dominant lineage among the highly-resistant *K. pneumoniae* isolates that we sequenced. ST147 was associated with both community- and the hospital-acquired infections, and with different infection sites, whereas other STs were predominantly uropathogens. Multiple antibiotic resistance and virulence determinants were detected in the genomes including extended-spectrum β-lactamases (ESBL) and carbapenemases. Many of these genes were plasmid-borne.

Conclusions. Our data suggest that the evolutionary success of ST147 may be linked with the acquisition of broad host-range plasmids, and their propensity to accrue AMR and virulence determinants. Although ST147 is a dominant lineage in many countries worldwide, it has not been previously reported as prevalent in Africa. Our data suggest an influx of new nosocomial pathogens with new virulence genes into African hospitals from other continents.

DATA SUMMARY

Klebsiella pneumoniae strains were isolated from clinical specimens, identified by bacteriological methods and tested for antibiotic resistance. Selected strains showing a broad spectrum of antibiotic resistance were sent to MicrobesNG

Received 13 June 2023; Accepted 08 January 2024; Published 22 January 2024
Author affiliations: [1]Department of Medical Laboratory Science, Masinde Muliro University of Science and Technology, Kakamega, Kenya; [2]Department of Medical Microbiology, Jomo Kenyatta University of Agriculture and Technology, Nairobi, Kenya; [3]Department of Biochemistry, Jomo Kenyatta University of Agriculture and Technology, Nairobi, Kenya; [4]Department of Medical Biochemistry, Masinde Muliro University of Science and Technology, Kakamega, Kenya; [5]Department of Biochemistry, University of Cambridge, Hopkins Building, Cambridge, UK; [6]Centre for Bioinformatics and Computational Biology, Department of Biochemistry, Genetics and Microbiology, University of Pretoria, Pretoria, South Africa.
***Correspondence:** Oleg N. Reva, oleg.reva@up.ac.za; Martin Welch, mw240@cam.ac.uk
Keywords: *Klebsiella pneumoniae*; environmental isolate; antibiotic resistance; whole genome sequence; MLST.
Abbreviations: AMC, amoxicillin/clavulanate; AMK, amikacin; AMR, antimicrobial resistance; BL, β-lactamase; CA, community-acquired; CAZ, ceftazidime; CIP, ciprofloxacin; CLSI, Clinical and Laboratory Standards Institute; COG, cluster of orthologous genes; CRO, ceftriaxone; CTX, cefotaxime; CXM, cefuroxime; EP, efflux pump; ESBL, extended-spectrum β-lactamase; F, female; FEP, cefepime; GI, genomic island; GM, gentamicin; HA, hospital-acquired; IMP, imipenem; IP, in-patient; IP/ICU, in-patient admitted in intensive care; KCGTRH, Kakamega County General Teaching and Referral Hospital; LMIC, low and middle income countries; M, male; MDR, multidrug-resistant; MEM, meropenem; MLST, multi-locus sequence typing; MPF, mate-pair formation; NIT, nitrofurantoin; OP, out-patient; P, pus; PBS, phosphate-buffered saline; PGAP, prokaryotic genome annotation pipeline; S, sputum; SAM, ampicillin/sulbactam; SDA, sabouraud dextrose agar; SNP, single nucleotide polymorphism; ST, sequence type; SXT, trimethoprim/sulfamethoxazole; T, tracheal aspirate; TZP, piperacillin/tazobactam; U, urine; UTI, urinary tract infection.
Whole genome assemblies of *Klebsiella pneumoniae* hospital isolates are at NCBI under the accession numbers listed in Table 1.
One supplementary figure and three supplementary tables are available with the online version of this article.
000667.v4 © 2024 The Authors

This is an open-access article distributed under the terms of the Creative Commons Attribution License.

(Birmingham, UK) for extraction of genomic DNA and sequencing on the Illumina HiSeq 2500 platform. DNA reads were quality controlled and assembled using SPAdes version 3.14.0. Contig annotation was performed by Prokaryotic Genome Annotation Pipeline (PGAP V.6.6). Antibiotic resistance genes were predicted using the RGI/CARD web-service (https://card.mcmaster.ca/analyze/rgi). MLST typing was performed and virulence genes were identified using Kleborate software and the VFDB database. Whole genome sequences were deposited at NCBI under accession numbers shown in Table 1.

INTRODUCTION

Klebsiella pneumoniae is emerging as a major clinical and public health threat, as is now reported to be responsible for up to one third of all Gram-negative infections. The organism is an opportunist and is becoming an increasingly prevalent source of nosocomial infections in the airways, urinary tract (where it ranks only second behind *Escherichia coli* as a causative agent) and surgical wound sites. *K. pneumoniae* is also a cause of serious community-acquired infections, especially pneumonias. The death rate for patients with *K. pneumoniae*-associated pneumonia is high, even following antibiotic treatment. Community-acquired infections appear to be linked with the spread of high-risk lineages such as ST147. Furthermore, clone ST147, is emerging globally as an important vehicle for the dissemination of AMR determinants [1]. In Low and Middle Income Countries (LMICs), multi-drug resistance (MDR) among clinically-significant Gram-negative bacteria is increasingly becoming a cause of increased morbidity, accounting for an estimated 40% of mortality [2].

Although community- and hospital-acquired transmission pathways for *K. pneumoniae* have clearly been demonstrated in high income countries [3], there is generally paucity of such evidenced studies from sub-Saharan Africa. Because of the absence of such data, policy makers and implementors therefore have limited evidence to rely on when allocating resources towards infection, prevention and control (IPC) programmes. In the current study, we attempt to rectify this and decipher the possible pathways that may have been involved in a sudden surge in *K. pneumoniae* occurrence in a regional referral health facility in Western Kenya during the period December 2015 to May 2016.

While monitoring for extended spectrum β-lactam resistance among *Enterobacteriaceae* isolates obtained from Kakamega County General Teaching and Referral Hospital (KCGTRH) in Western Kenya, we noticed an increase in the frequency of multi-drug resistant *K. pneumoniae*. These isolates were resistant to a broad range of structurally distinct classes of antimicrobial agent. We therefore used whole genome sequencing (WGS) to analyse the genetic structure of a selection of these *K. pneumoniae* isolates. Moreover, by analysing the encoded virulent traits, antibiotic resistance genes, and plasmid sequences in each isolate, we provide an evidence-based description of the local species diversity associated with the outbreak, as well as possible transmission pathways.

Our study strongly suggests that antibiotic resistance is acquired in the *K. pneumoniae* lineages through acquisition of virulence plasmids that are enriched with antibiotic/antimicrobial resistance genes (ARGs). We show that WGS is a powerful, and increasingly economical tool in the fight against communicable infections in the LMIC healthcare envioronment, and can even potentially discriminate between the lineages that cause community-associated infections and nosocomial infections. Such data should be invaluable for health authorities in LMIC, enabling them to identify key areas for intervention and resource deployment.

METHODS

Bacterial isolation and identification

Samples were obtained from KCGTRH (Kenya) over the period from December 2015 to May 2016. *K. pneumoniae* were recovered from patients seeking treatment for various ailments including respiratory tract infections, urinary tract infections (UTIs), sepsis and wounds. Infections were designated as hospital- or community-acquired to distinguish between different possible sources of infection. The nosocomial infections were defined as those acquired by patients after hospitalization, manifesting 48 h after admission. For the urine specimens, midstream urine samples were collected and cultured on MacConkey agar (Hi-Media, India). Wound samples were collected by swabbing the wound, and were also cultured on MacConkey agar (Hi-Media, India). For blood culture, two sets of sterile draws (aerobic and anaerobic) from two separate venipuncture sites were drawn and the samples were then incubated in a blood culture incubator (BACTEC) at 35–37°C for 24–36 h. The blood samples were then sub-cultured inoculated by streaking on blood agar base BAP, Chocolate blood agar, MacConkey and Sabouraud Dextrose Agar (SDA) for further identification. Colony-pure isolates were confirmed as *K. pneumoniae* using API20E (BioMérieux) biochemical test strips following the manufacturer's instructions.

Antimicrobial susceptibility testing

A total of 27 *K. pneumoniae* isolates were subjected to antibiotic susceptibility profiling using the Kirby Bauer disc diffusion method. We tested amikacin, amoxicillin/clavulanate, ampicillin/sulbactam, cefepime, cefotaxime, ceftazidime, ceftriaxone, cefuroxime, gentamicin, imipenem, meropenem, nitrofurantoin, piperacillin/tazobactam, ciprofloxacin, and trimethoprim/sulfamethoxazole. *Escherichia coli* ATCC 25922 was used as a reference as per Clinical and Laboratory Standards Institute (CLSI) guidelines (2017). Of the 27 isolates tested, nine were multi-drug resistant. For long-term storage, cultures of the *K. pneumoniae* isolates were frozen at −80°C in trypticase soy broth supplemented with 15% v/v glycerol.

Genomic DNA sequencing

Whole genome sequencing was carried out by MicrobesNG (Birmingham, UK) using an Illumina HiSeq 2500 platform. Briefly, a single colony of each strain was picked and suspended in 100 µl of sterile 1×phosphate-buffered saline (PBS) (Oxoid, UK). The suspension was spread thickly (using a sterile loop) onto a fresh LB-agar plate and incubated at 37°C overnight. Dense colony growth was then scraped off and sent to MicrobesNG in supplied bar-coded bead tubes. Sequencing (30-fold depth) was carried out using an Illumina HiSeq 2500 platform, with 2×250 bp paired-end reads. The reads were trimmed using Trimmomatic v0.30 with a sliding window quality cut-off of Q15. The *de novo* assembly of contigs was done using SPAdes version 3.14.0 with default settings. The resulting contigs were scaffoled by alignment to the closest reference sequences found in NCBI by MegaBLAST search. The closest reference genomes are listed in Table 1. Gaps between contigs were patched by the respective genomic fragments from the reference genomes. Then the original DNA reads were mapped against the resulting genome sequences by the Bowtie2 algorithm implemented in Unipro UGENE v48.1 [4] to verify the patched regions and generate consensus sequences. Preliminary genome annotation was done using the RAST Annotation Server [5] with automatic fix error function, which checks and fixes possible sequencing and genome assembly errors. The NCBI Prokaryotic Genome Annotation Pipeline (PGAP v.6.6) [6] was later used to refine the annotation. BioPython v1.81 was used to find *dnaA* coding sequences in the annotated genomes and to rearrange the sequences in a way such that each sequence starts 400 bp upstream of *dnaA* in the leading DNA strand (i.e. to correspond to the chromosomal replication origin). The genomic sequences were finalized by another round of mapping the original DNA reads against the genome sequences to confirm circularity of the chromosomes and to verify modifications introduced by the RAST fix error function. Quality of the resulting consensus sequences was controlled by the programme CheckM2 v1.0.2 [7]. Plasmid contigs were identified using mlplasmids 1.0.0 [8] and the *mob_recon* function of the Mob-Suite v3.1.7 utility (https://github.com/phac-nml/mob-suite) [9] with the default parameter setting. Further analysis and genotyping of the plasmid sequences was performed by using function *mob_typer* of Mob-Suite. Whole genome sequences were deposited at NCBI under the BioSample accession numbers shown in Table 1.

Gene ortholog prediction and phylogenetic inferencing

Clusters of orthologous groups (COGs) in sequenced genomes were predicted using the programme OrthoFinder [10] with default parameters. The sequences of each COG were aligned using the MUSCLE v3.6 algorithm with the parameters set by default [11]. Ambiguous parts of the alignments were removed using the programme Gblocks v0.91 [12] with the default parameter settings. COG alignments were concatenated using BioPython scripts into superstring alignments for further phylogenetic inferences using the Neighbour-Joining algorithm implemented MEGA-X [13]. Alignment of plasmid sequences was performed by Mauve 20150226 [14]. Here, the sequence similarity is the ratio of base pairs shared between two genomes to their average genome length. This

Table 1. Accession numbers of *K. pneumoniae* genome sequences at GenBank NCBI

Strain no.	BioSample	CheckM2 metrics (Completeness / contamination)	Reference genomes
CK1	SAMN24338476	100% / 0.76%	CP054990.1
CK2	SAMN31566495	99% / 0.2%	CP035196.1
CK3	SAMN31566649	100% / 0.65%	CP044047.1
CK4	SAMN24369220	100% / 0.54%	CP023839.1
CK5	SAMN24369239	99.9% / 0.95%	CP050371.1
CK6	SAMN24369281	100% / 1.52%	CP024916.1
CK7	SAMN24369410	100% / 1.52%	CP024916.1
CK8	SAMN24370162	100% / 1.12%	CP044047.1
CK9	SAMN24371033	100% / 1.52%	CP024916.1
K10	SAMN24371047	100% / 1.1%	CP052224.1
K11	SAMN24377653	100% / 1.1%	CP052224.1
K12	SAMN24377668	100% / 1.1%	CP052224.1
K13	SAMN24386524	100% / 1.1%	CP052286.1
K14	SAMN24386561	100% / 0.08%	CP043859.1
K15	SAMN24386923	100% / 1.1%	CP030315.1

similarity estimate is then converted to a distance value for the NJ distance matrix. The progressiveMauve algorithm produces a dendrogram of sequence similarity relations stored in an alignment.guide_tree file. The dendrogram was visualized by MEGA-X.

Genome typing, virulence and antibiotic resistance prediction

Species confirmation of the sequenced microorganisms, multilocus sequence typing (MLST), capsule typing, antibiotic resistance and virulence predictions were done using Kleborate v2.3.2 (https://github.com/klebgenomics/Kleborate) [15] and Kaptive v2.0.7 (https://github.com/klebgenomics/Kaptive) [16]. MLST predictions were controlled by searching through the BIGSdb (https://bigsdb.pasteur.fr/) and CBS databases [17, 18]. Antibiotic resistance and virulence genes additionally were predicted using Resistance Gene Identifier (RGI 6.0.3) with the CARD database (https://card.mcmaster.ca/analyze/rgi) [19] and VFDB database records for *Klebsiella* [20]. The Kleborate and Kaptive predictions were combined with the phylogenetic tree and visualized using the Microreact Web-service (https://microreact.org/). Mobile horizontally-acquired genomic islands (GIs) were identified using SeqWord Genomic Island Sniffer (http://seqword.bi.up.ac.za/sniffer/index.html) [21]. SeqWord identifies horizontally-acquired genomic elements by analysing patterns of oligonucleotide signatures.

RESULTS

Antibiotic susceptibility profiles of the selected *Klebsiella pneumoniae* isolates

Hospital- and community-acquired *Klebsiella* pathogens were isolated from a range of infection types and sites over a ca. 6 month period, as outlined in the Methods. Each isolate was tested for its ability to grow on a selection of structurally-diverse antibiotics with different targets and different modes of action. The antibiotics tested included β-lactams, cephalosporins and β-lactamase inhibitors (amoxicillin/clavulanate (AMC), piperacillin/tazobactam (TZP), ceftazidime (CAZ), meropenem (MEM), cefotaxime (CTX), cefuroxime (CXM), ceftriaxone (CRO), cefepime (FEP), ampicillin/sulbactam (SAM) and imipenem (IMP), aminoglycosides (gentamicin [GM] and amikacin [AMK]), a DNA gyrase inhibitor (ciprofloxacin, CIP), folate metabolism inhibitors (trimethoprim/sulfamethoxazole, SXT), and nitrofurantoin (NIT), which is known to be a very effective antibiotic with a poorly-defined mode of action. Of the 27 isolates we tested, nine were resistant to all of these antibiotics. The genome of each of these super-resistant isolates was sequenced. Some isolates were resistant to all of the tested antibiotics except one; for example, MEM (isolate CK1), IMP (isolate CK3), SXT (isolate CK4), NIT (isolate CK5) and AMK (isolate K15). We also sequenced the genome of each of these isolates too. Finally, isolate CK2 was also included in the genome sequencing, since it was sensitive to all of the antibiotics tested except SAM. With the exception of CK1, all of the isolates subjected to whole genome sequencing were harvested in the period March 2016–April 2016. Genomic DNA was collected from single colonies of the sampled *K. pneumoniae* strains following overnight growth on LB-agar at 37°C. DNA samples were sequenced using an Illumina HiSeq 2500 platform with 2×250 bp paired-end reads, and assembled using SPAdes v3.14.0. The final genome assemblies were annotated by PGAP. Genome completeness and purity were confirmed by CheckM2 analysis (Table 1). All of the isolates were unambiguously identified as *Klebsiella pneumoniae* following analysis using of the whole genome sequences using Kleborate (Table S2, available in the online version of this article).

Epidemiology and typing of *Klebsiella pneumoniae* isolates

The obtained genome sequences were used for MLST typing of the isolates. Nine out of the 15 isolates were assigned to ST147 (including a sub-type ST147-1LV, where 1LV indicates that an additional SNP is present compared with the previously published ST). The ST147 isolates were split in the phylogenetic tree into several clusters characterized by two different polysaccharide capsule types, KL64 and KL10. Two isolates CK3 and CK8 belonged to ST231 and ST231-1LV (respectively) but had the same capsule type, KL51. The remaining four minor isolates were assigned to ST11, ST14, ST1634 and ST1801, and were characterized as having three different capsule types (Fig. 1). Additional information about sources and dates of isolation of the clinical isolates, and their respective STs, are given in Table 2.

Phylogenetic relations between *Klebsiella pneumoniae* isolates

DNA reads generated by the Illumina sequencer were assembled using SPAdes into contigs, which then were scaffolded and joined to yield whole genome sequences. The genomes were annotated using the NCBI genome annotation robot PGAP. Following this, clusters of orthologous protein-coding genes (COG) were identified, translated to amino acid sequences, and aligned. Alignments of protein sequences of all COGs (2637 orthologous protein-coding genes) were concatenated into a superstring alignment comprising 805256 positions (inclusive of amino acid residues and gaps). A Neighbour-Joining phylogenetic tree was then generated based on this alignment. The genetic typing by MLST is in good agreement with the phylogenetic relations inferred between the isolates based on the NJ tree of the superstring alignment (Fig. 1). The phylogenetic tree structure was consistent with the genomic MLST typing. The isolates of ST147 were associated with both hospital-acquired (HA) and community-acquired (CA) infections, and with a variety of infection sites. By contrast, isolates CK1, CK2, CK3, CK8, and CK14 all belonged to different STs, but were all associated with urinary tract infections.

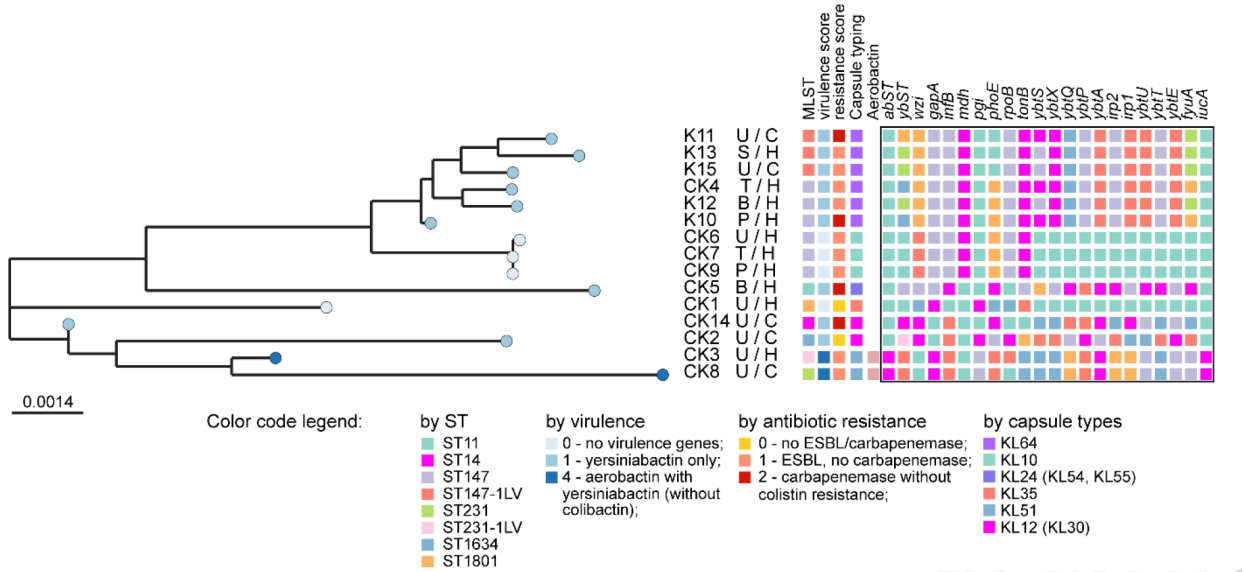

Fig. 1. Phylogeny and genotypes of the *K. pneumoniae* isolates shown in the Neighbour-Joining tree as nodes shaded in accordance with Kleborate predicted virulence scores. Predicted MLST types of the isolates (ST), extracellular capsule types, virulence and antibiotic resistance categories are depicted by colour as explained in the figure legend. Sequence variants of marker genes traditionally used for MLST are depicted in the framed box. Infection sites and infection acquisition were abbreviated: Pus (p); Blood (b); Urine (u); Sputum (s); Tracheal aspirate (t); Hospital-acquired (h); Community-acquired (c).

Identification of antibiotic resistance genes and their distribution between chromosomes, genomic islands and plasmids

Numerous ARG and virulence determinants were identified in the isolates using the public databases RGI-CARD, VFDB and Kleborate software (Table S1). These included various efflux pumps (*acrBDFS*, *emrABKRY*, *oqxAB*, *tet(E)*, *mdfABCEFNOP*) and their known regulators (*baeR*, *crp*, *evgA* and *gadX*), as well as a selection of other important regulators that have been previously

Table 2. Source and MLST typing of the sequenced *K. pneumoniae* isolates

Isolate identity	Sequence type (ST)	Sample site	Type of hospital care	Date of isolation	Age	Sex	Hospital (H)/Community (C) aquired
CK1	1801	U	IP	28 December 2015	42	M	H
CK2	1634	U	OP	21 March 2016	66	F	C
CK3	231	U	IP*	07 March 2016	36	F	H
CK8		U	OP	01 April 2016	62	F	C
CK5	11	B	IP	11 March 2016	39	F	H
CK14	14	U	OP	18 March 2016	36	M	C
CK4	147	T	IP/ICU*	25 April 2016	59	F	H
CK6		U	IP/ICU	25 April 2016	48	M	H
CK7		T	IP	07 May 2016	72	F	H
CK9		P	IP	05 April 2016	28	M	H
K10		P	IP*	21 March 2016	56	M	H
K11		U	OP	30 March 2016	49	F	C
K12		B	IP/ICU*	07 April 2016	57	F	H
K13		S	IP*	31 March 2016	11	F	H
K15		U	OP	11 March 2016	47	F	C

B, Blood; F, Female; IP*, In-patient, but admitted for <48 h; IP, In-patient; IP/ICU, In-patient admitted in intensive care; M, Male; OP, Out patient; P, Pus; S, Sputum ; T, Tracheal aspirate; U, Urine.

associated with antibiotic resistance (*baeSR*, *evgAS* and *kdpE*). We also confirmed the presence of certain stress response genes (*cpxA*, *marA* and *tolC*) that have been implicated in modulating drug efflux under some conditions [22, 23]. Three β-lactamases were identified; the sulfhydryl reagent-variable β-lactamase SHV-52, the extended spectrum β-lactamase CTX-M-130, and the AmpC-family enzyme [24], FOX-1. Additional AMR-associated genes included the MFS-family drug transporters, *mdtH* and *mdtM*, and an SMR-type efflux pump (*kpnEF*). Moreover, several genes involved in modifying the bacterial cell envelope in response to stress were identified, including genes involved in lipid A biosynthesis (*pmrF* and *ugd*) and modification (*eptA*) [25].

Other AMR-associated genes were also present in the isolates. These included *fosA5* (a glutathione *S*-transferase implicated in fosfomycin resistance), *sul3* (implicated in sulfonamide resistance), chloramphenicol exporters (*cmlB* and *floR*) and chloramphenicol acetyltransferases (*catI* and *catB11*), tetracycline resistance genes (including *tet*(*32*) which confers ribosomal protection against tetracycline, and the tigecycline resistance gene, *tet(X4)*), a phosphotransferase (*mphB*) that phosphorylates macrolide antibiotics, the bacitracin phosphorylase *bacA*, and the microcin exporter *yojI*.

Many of the above mentioned ARGs were distributed across the core parts of the chromosomes (conserved across the isolates) and various mobile elements. For example, *acrE*, *evgA*, *evgS* and *ugd* were found within predicted horizontally-acquired genomic islands (GIs) in all the isolates. By contrast, many of the other AMR determinants exhibited a less homogenous distribution between isolates. For example, the β-lactamase CTX-M-130, fosfomycin resistance gene *fosA5*, chloramphenicol exporter *cmlB*, efflux pump regulator *acrS*, MFS transporter *mdtM*, tetracycline resistance gene *tet32*, inner membrane transporter *acrFE*, efflux pumps *oqxAB* and *mdfA*, antibiotic resistance genes *yojI* and *floR*, and the multi-drug transporter *ymrKY* were each found within GIs in at least one of the selected isolates. The numbers of ARG associated with GIs in the selected microorganisms varied from three (in CK14) to seven (in CK5 and CK8).

An overall prediction of the virulence and antibiotic resistance status of the isolates was performed using Kleborate (Fig. 1 and Table S2). Two strains, CK3 and CK8 (both ST231) isolated from the hospital and from the community (respectively), were predicted to encode both the aerobactin and yersiniabactin iron acquisition systems. These siderophore systems contribute significantly to the virulence of *Klebsiella* and *Escherichia* spp. [26–28]. Another nine strains possessed the yersiniabactin system alone, and four of the isolates encoded neither siderophore system. All strains except for two (CK1 and CK2) encoded extended-spectrum β-lactamases (ESBL, CTX-M-15) [29] and Kleborate predicted carbapenemases in four strains (OXA-181 in strains K10, K11 and CK14; and NDM-1 in CK5) rendering resistance to carbapenem antibiotics [30]. The streptomycin resistance genes, *strAB*, and the sulfonamide resistant dihydropteroate synthase, *sul2*, were found in CK5, K10, K11, K12, K13, and K15. Strain K14 also possessed an aminoglycoside-modifying enzyme, encoded by *sat*-2. Mutations in *gyrA* and *parC* associated with fluoroquinolone resistance [31] were identified in all strains except for CK1 and CK2. Finally, CK3, CK4, CK8, K10, K11, K12, K13, K14 and K15 all carried mutations in the outer membrane porins, *ompK35* and *ompK36*, associated with doripenem-colistin resistance [32].

Plasmid-associated contigs were identified using mlplasmid (a support vetor machine model trained on *K. pneumoniae*). Alignment of plasmid contigs against each other using the programme Mauve (Fig. S1) allowed us to construct the dendrogram in Fig. 2. A search for similar plasmids in the NCBI was perfomed using Mob-Suite software [9]. All the plasmids were predicted by this tool as conjugative broad-range plasmids belonging to mate-pair formation (MPF) types I and F. The MPF classification of the plasmids is in agreement with their clustering in the dendrogram (Fig. 2). The full list of determined Mob-Suite metrics is shown in Table S3. In parallel, we also screened for plasmid-borne ARGs. A number of plasmid-associated ARGs were identified, including β-lactamases (BL) and efflux pumps (EP) (Fig. 2). Interestingly, many of these ARGs were paralogues of genes already present on the chromosomes. This may indicate that plasmid-borne ARGs (and possibly, also those associated with other horizontally-acquired GIs) may confer a greater degree of antibiotic resistance than the homologous genes located in the core genome. Alternatively, the elevated copy number of the duplicated genes may be responsible (i.e. a gene dosage effect). Detailed information on the specific plasmid-borne ARGs associated with each strain is provided in Supplementary Table S2.

Alignment of the plasmid contigs against one another revealed some level of sequence similarity, with the greatest identity (> 90%) for the plasmids from the ST147 strains K10, K12, K13 and K15, all of which belong to ST147. We also note that the hospital-acquired ST147 strains CK6, CK7 and CK9 all appear to share common plasmids.

However, this similarity between plasmids only to some extent corresponded with the phylogenetic relationship between the isolates (cf. Fig. 1 versus Fig. 2). Two types of plasmid were associated with the isolates of ST147. Isolates K10, K12, K13 and K15 contain large plasmids of 440–490 kb size. Assuming that the evolution of virulence plasmids in the antibiotic era is towards rendering multidrug resistance, these plasmids may have evolved from the smaller, ancestral plasmids represented by those in isolates CK2 (ST1634) and CK4 (ST147) through the acquisition of additional ARGs. The ST147 isolates CK6, CK7 and CK9 contain another type of plasmid (size range 301–335 kb). These plasmids may also have evolved from smaller plasmids, although here, the ancestor appears to be similar to the plasmid in isolate CK1 (ST1801). This evolutionary pathway too is associated with an increase in the number of ARGs; indeed, the plasmids from CK6, CK7 and CK9 contain the largest number of plasmid-born ARG seen in the current study, even outnumbering the ARGs present in the much larger plasmids from isolates K10, K12, K13 and K15 (Fig. 2). The plasmid from the antibiotic-susceptible isolate, CK2, carried the smallest number of ARGs.

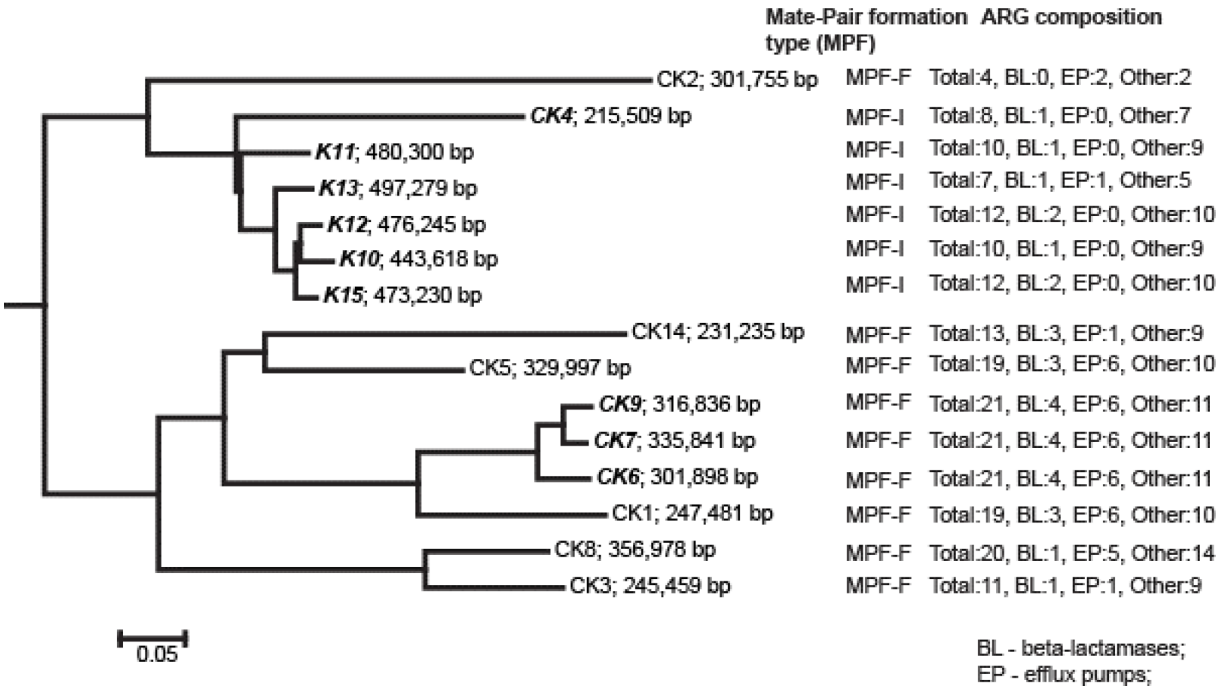

Fig. 2. Dendrogram showing sequence similarity of the plasmids found in the *K. pneumoniae* isolates. The tree was rooted at the average branch location not representing the common ancestry. The total numbers of antimicrobial resistance genes (ARG) and the numbers of β-lactamases (BL), efflux pumps (EP) and other categories of ARG are shown on the right-hand side of the figure.

DISCUSSION

Our data indicate that ST147 was enriched among the highly-resistant *K. pneumoniae* isolates that we selected for whole genome sequencing. Indeed, of the 15 isolates that we analysed this way, nine were ST147. On the other hand, according to the Kleborate predictions (Fig. 1), the most worrisome isolates were the two strains of ST231.

To date, *K. pneumoniae* ST147 has been reported as a causative agent of high-risk hospital infections around the world, but mostly in the northern hemisphere [33–36]. However, at the time of publication, KCGTRH receives 5–10 patients per day with *K. pneumoniae*-associated infections. Our data indicate that (unless KCGTRH is somehow unusual) ST147 is now common among the hospital isolates from Western Kenya. By contrast, other recent studies did not observe *K. pneumoniae* ST147 in African hospitals [37–39], and *K. pneumoniae* lineages have been reported to be diverse and dominated by ST131, ST335, ST1193, ST10, ST14, ST15 and ST307. Indeed, only one paper has mentioned multi-drug resistant isolates of *K. pneumoniae* ST147 from Kenya [38]. However, although that paper was published in 2022, we note that sampling of the isolates analysed therein happened between 2015–2020. Given the dominance of ST147 observed in the current study, this may suggest that ST147 emerged in the region around that time. The significant shift in population structure of hospital-associated *K. pneumoniae* from the diverse lineages in [37] and [38] to the dominance of ST147 reported here is curious. One possibility is that the shift is linked with human migration, or by the abuse of antibiotics in the region (we recall here the plasmid-borne accrual of AMR and virulence genes in the ST147 isolates). Another possible factor may be that this lineage is not confined by an infection site. Consistent with this, there was no obvious correlation between the prevalence of ST147 and the type of infection (wound exudate, blood, tracheal aspirate, expectorated sputum, urinary). Moreover, we note that ST147 strains could equally be isolated either from hospitals or from the community [38]. Finally, another formal possibility is that discrepencies in the reported frequency of ST147 simply reflect a somewhat uneven geographical distribution of the lineage between clinics, or a recent influx of the lineage from other parts of the world. If so, this is of major concern, especially given that ST147 provides a formidable platform for the dissemination of AMR. At the very least, a comprehensive effort will be needed to understand the true distribution of this clone-type in the east African region.

Antibiotic resistance genes were identified in both the core genome and in genomic islands of all the isolates we examined in this study. We also identified many paralogous resistance determinants in the plasmidome of each isolate. The evolutionary success of ST147 appears to be associated with the acquisition of two broad-range plasmids, each enriched in ARGs. The proclivity of *K. pneumoniae* lineages to acquire plasmids from one another is well known [33]. However, AMR may not be the only determinant of ST success in the region. ST231 was also represented in our dataset, and worryingly, not only manifested multi-drug resistance, but

was also enriched in virulence determinants as well as ARGs. These features were encoded by-and-large, on horizontally-acquired GI insertions in the ST231 chromosomes. Isolates of ST231 from China, Türkiye and Africa have previously been characterized and found to be associated with broad spectrum antibiotic resistance, especially against carbapenem antibiotics [40, 41].

Funding information
This work was generously supported by a grant from the Cambridge-Africa Initiative (Alborada Trust) to AWS, ONR and MW. TJOB was supported by a PhD studentship from the National Centre for Replacement, Refinement and Reduction of Animals in Research (NC3Rs).

Author contributions
Conceptualization: A.W.S., M.W.; Funding Acquisition: A.W.S., M.W., O.N.R.; Investigation: V.D., A.N.K., D.K., A.W.S., T.J.O'B., M.W., O.N.R; Data curation: A.W.S., M.W., O.N.R.; Project administration: A.W.S., M.W.; Software: O.N.R.; Supervision: A.W.S., M.W.; Validation: M.W., O.N.R.; Visualization: O.N.R.; Writing - Original Draft: A.W.S., M.W., O.N.R.; Writing - review & editing: V.D., A.N.K., D.K., A.W.S., T.J.O'B., M.W., O.N.R.

Conflicts of interest
The authors declare that the research was conducted in the absence of any relationships that could be construed as a potential conflict of interest.

Ethical statement
Ethical approval was given by Jaramogi Oginga Odinga Referral and Teaching Hospital IREC reference number ERC.IB/VOL.1/141, and a letter for permission to collect data was provided by Kakamega County Referral Hospital IREC.

References

1. Pitout JDD, Peirano G, Kock MM, Strydom KA, Matsumura Y. The Global ascendency of OXA-48-type carbapenemases. *Clin Microbiol Rev* 2019;33:e00102-19.
2. Founou RC, Founou LL, Essack SY. Clinical and economic impact of antibiotic resistance in developing countries: a systematic review and meta-analysis. *PLoS One* 2017;12:e0189621.
3. Martin RM, Bachman MA. Colonization, infection, and the accessory genome of *Klebsiella pneumoniae Front Cell Infect Microbiol* 2018;8:4.
4. Okonechnikov K, Golosova O, Fursov M, UGENE Team. Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics* 2012;28:1166–1167.
5. Brettin T, Davis JJ, Disz T, Edwards RA, Gerdes S, *et al*. RASTtk: a modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. *Sci Rep* 2015;5:8365.
6. Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, *et al*. NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Res* 2016;44:6614–6624.
7. Chklovski A, Parks DH, Woodcroft BJ, Tyson GW. CheckM2: a rapid, scalable and accurate tool for assessing microbial genome quality using machine learning. *Nat Methods* 2023;20:1203–1212.
8. Arredondo-Alonso S, Rogers MRC, Braat JC, Verschuuren TD, Top J, *et al*. mlplasmids: a user-friendly tool to predict plasmid- and chromosome-derived sequences for single species. *Microb Genom* 2018;4:11.
9. Robertson J, Nash JHE. MOB-suite: software tools for clustering, reconstruction and typing of plasmids from draft assemblies. *Microb Genom* 2018;4:e000206.
10. Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol* 2019;20:238.
11. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 2004;32:1792–1797.
12. Talavera G, Castresana J. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst Biol* 2007;56:564–577.
13. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. *Mol Biol Evol* 2018;35:1547–1549.
14. Darling ACE, Mau B, Blattner FR, Perna NT. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res* 2004;14:1394–1403.
15. Lam MMC, Wick RR, Watts SC, Cerdeira LT, Wyres KL, *et al*. A genomic surveillance framework and genotyping tool for *Klebsiella pneumoniae* and its related species complex. *Nat Commun* 2021;12:4188.
16. Wyres KL, Wick RR, Gorrie C, Jenney A, Follador R, *et al*. Identification of *Klebsiella* capsule synthesis loci from whole genome data. *Microb Genom* 2016;2:e000102.
17. Jolley KA, Maiden MCJ. BIGSdb: scalable analysis of bacterial genome variation at the population level. *BMC Bioinformatics* 2010;11:595.
18. Larsen MV, Cosentino S, Rasmussen S, Friis C, Hasman H, *et al*. Multilocus sequence typing of total-genome-sequenced bacteria. *J Clin Microbiol* 2012;50:1355–1361.
19. Alcock BP, Raphenya AR, Lau TTY, Tsang KK, Bouchard M, *et al*. CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database. *Nucleic Acids Res* 2020;48:D517–D525.
20. Chen L, Zheng D, Liu B, Yang J, Jin Q. VFDB 2016: hierarchical and refined dataset for big data analysis--10 years on. *Nucleic Acids Res* 2016;44:D694–7.
21. Bezuidt O, Lima-Mendez G, Reva ON. Seqword gene Island Sniffer: a program to study the lateral genetic exchange among bacteria. *World Acad Sci Eng Technol* 2009;22:1169–1174.
22. Nishino K, Nikaido E, Yamaguchi A. Regulation and physiological function of multidrug efflux pumps in *Escherichia coli* and *Salmonella*. *Biochim Biophys Acta* 2009;1794:834–843.
23. Wang H, Zhang H, Zhang H, Jin J, Xie Y. Outer membrane channel protein TolC regulates *Escherichia coli* K12 sensitivity to plantaricin BM-1 via the CpxR/CpxA two-component regulatory system. *Probiotics Antimicro Prot* 2021;13:238–248.
24. Moland ES, Black JA, Ourada J, Reisbig MD, Hanson ND, *et al*. Occurrence of newer beta-lactamases in *Klebsiella pneumoniae* isolates from 24 U.S. hospitals. *Antimicrob Agents Chemother* 2002;46:3837–3842.
25. James CE, Mahendran KR, Molitor A, Bolla JM, Bessonov AN, *et al*. How beta-lactam antibiotics enter bacteria: a dialogue with the porins. *PLoS One* 2009;4:e5453.
26. Lawlor MS, O'connor C, Miller VL. Yersiniabactin is a virulence factor for *Klebsiella pneumoniae* during pulmonary infection. *Infect Immun* 2007;75:1463–1472.
27. Gao Q, Wang X, Xu H, Xu Y, Ling J, *et al*. Roles of iron acquisition systems in virulence of extraintestinal pathogenic *Escherichia coli*: salmochelin and aerobactin contribute more to virulence than heme in a chicken infection model. *BMC Microbiol* 2012;12:1–2.
28. Russo TA, Olson R, Macdonald U, Metzger D, Maltese LM, *et al*. Aerobactin mediates virulence and accounts for increased siderophore production under iron-limiting conditions by hypervirulent (hypermucoviscous) Klebsiella pneumoniae. *Infect Immun* 2014;82:2356–2367.
29. Paterson DL, Hujer KM, Hujer AM, Yeiser B, Bonomo MD, *et al*. Extended-spectrum beta-lactamases in *Klebsiella pneumoniae* bloodstream isolates from seven countries: dominance and

widespread prevalence of SHV- and CTX-M-type beta-lactamases. *Antimicrob Agents Chemother* 2003;47:3554–3560.

30. Pitout JDD, Nordmann P, Poirel L. Carbapenemase-producing *Klebsiella pneumoniae*, a key pathogen set for global nosocomial dominance. *Antimicrob Agents Chemother* 2015;59:5873–5884.
31. Kareem SM, Al-Kadmy IMS, Kazaal SS, Mohammed Ali AN, Aziz SN, *et al*. Detection of gyrA and parC mutations and prevalence of plasmid-mediated quinolone resistance genes in *Klebsiella pneumoniae Infect Drug Resist* 2021;14:555–563.
32. Clancy CJ, Chen L, Hong JH, Cheng S, Hao B, *et al*. Mutations of the ompK36 porin gene and promoter impact responses of sequence type 258, KPC-2-producing *Klebsiella pneumoniae* strains to doripenem and doripenem-colistin. *Antimicrob Agents Chemother* 2013;57:5258–5265.
33. Guo L, An J, Ma Y, Ye L, Luo Y, *et al*. Nosocomial outbreak of OXA-48-producing *Klebsiella pneumoniae* in a Chinese hospital: clonal transmission of ST147 and ST383. *PLoS One* 2016;11:e0160754.
34. Falcone M, Giordano C, Barnini S, Tiseo G, Leonildi A, *et al*. Extremely drug-resistant NDM-9-producing ST147 *Klebsiella pneumoniae* causing infections in Italy, May 2020. *Euro Surveill* 2020;25:48.
35. Protonotariou E, Poulou A, Politi L, Sgouropoulos I, Metallidis S, *et al*. Hospital outbreak due to a *Klebsiella pneumoniae* ST147 clonal strain co-producing KPC-2 and VIM-1 carbapenemases in a tertiary teaching hospital in Northern Greece. *Int J Antimicrob Agents* 2018;52:331–337.
36. Peirano G, Chen L, Kreiswirth BN, Pitout JDD. Emerging antimicrobial-resistant high-risk *Klebsiella pneumoniae* clones ST307 and ST147. *Antimicrob Agents Chemother* 2020;64:e01148-20.
37. Decano AG, Pettigrew K, Sabiiti W, Sloan DJ, Neema S, *et al*. Pan-resistome characterization of uropathogenic *Escherichia coli* and *Klebsiella pneumoniae* strains circulating in Uganda and Kenya, isolated from 2017-2018. *Antibiotics* 2021;10:12.
38. Muraya A, Kyany'a C, Kiyaga S, Smith HJ, Kibet C, *et al*. Antimicrobial resistance and virulence characteristics of *Klebsiella pneumoniae* isolates in Kenya by whole-genome sequencing. *Pathogens* 2022;11:545.
39. Musila L, Kyany'a C, Maybank R, Stam J, Oundo V, *et al*. Detection of diverse carbapenem and multidrug resistance genes and high-risk strain types among carbapenem non-susceptible clinical isolates of target gram-negative bacteria in Kenya. *PLoS One* 2021;16:e0246937.
40. Tian D, Pan F, Wang C, Sun Y, Zhang H. Resistance phenotype and clinical molecular epidemiology of carbapenem-resistant *Klebsiella pneumoniae* among pediatric patients in Shanghai. *Infect Drug Resist* 2018;11:1935–1943.
41. Spadar A, Perdigão J, Campino S, Clark TG. Large-scale genomic analysis of global *Klebsiella pneumoniae* plasmids reveals multiple simultaneous clusters of carbapenem-resistant hypervirulent strains. *Genome Med* 2023;15:3.

The Microbiology Society is a membership charity and not-for-profit publisher.

Your submissions to our titles support the community – ensuring that we continue to provide events, grants and professional development for microbiologists at all career stages.

Find out more and submit your article at microbiologyresearch.org

Peer review history

VERSION 3

Editor recommendation and comments

https://doi.org/10.1099/acmi.0.000667.v3.3
© 2024 Munnoch J. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License.

John Munnoch; University of Strathclyde, SIPBS, UNITED KINGDOM, Glasgow

Date report received: 08 January 2024
Recommendation: Accept

Comments: This study would be a valuable contribution to the existing literature.

SciScore report

https://doi.org/10.1099/acmi.0.000667.v3.1
© 2023 The Authors. This is an open-access article report distributed under the terms of the Creative Commons License.

iThenticate report

https://doi.org/10.1099/acmi.0.000667.v3.2
© 2023 The Authors. This is an open-access article report distributed under the terms of the Creative Commons License.

Author response to reviewers to Version 2

Response to Reviewers (Revision 2)

Editor comments:

Thank you for your revisions, reviewers were happy with your responses, there are only now minor changes required (for example some software versions). No additional data processing is required, however where possible please include the information requested.

DONE

Reviewers' comments and responses to custom questions:

Please rate the manuscript for methodological rigour

Reviewer 1: Very good

Please rate the quality of the presentation and structure of the manuscript

Reviewer 1: Good

To what extent are the conclusions supported by the data?

Reviewer 1: Strongly support

Do you have any concerns of possible image manipulation, plagiarism or any other unethical practices?

Reviewer 1: No:

If this manuscript involves human and/or animal work, have the subjects been treated in an ethical manner and the authors complied with the appropriate guidelines?

Reviewer 1: Yes:

Reviewer 1 Comments to Author: I am happy that the authors have sufficiently addressed all my comments from the previous round of review. Their work contributes to the surveillance of an important pathogen in an under-reported region.

Software version numbers for CheckM2, Kleborate and Kaptive are missing and should be included.

DONE

Please rate the manuscript for methodological rigour

Reviewer 2: Good

Please rate the quality of the presentation and structure of the manuscript

Reviewer 2: Satisfactory

To what extent are the conclusions supported by the data?

Reviewer 2: Partially support

Do you have any concerns of possible image manipulation, plagiarism or any other unethical practices?

Reviewer 2: No:

If this manuscript involves human and/or animal work, have the subjects been treated in an ethical manner and the authors complied with the appropriate guidelines?

Reviewer 2: Yes:

Reviewer 2 Comments to Author: 1. Methodological rigour, reproducibility and availability of underlying data

The Methods section is now written clear and the reanalysis using Kleborate suggested by reviewer one adds valuable information to the manuscript.

* The information of how many isolates were tested for AMR in total and what is the

percentage of multi-drug resistant isolates is missing in the method section. This is only

stated in the Results section.

DONE

The information is on Ln122. Given that the % resistant isolates is a result, rather than a method, we

would prefer to leave this where it currently is (Ln200).

It would be beneficial to know how many K. pneumonia strains were isolated during the sampling time and how many of those are multi drug resistant. The authors stated that 5-10 patients a day present with K. pneumonia currently. Is that 2023? Is there a change in the presents of multidrug resistance K. pneumonia? If there is an increase from the sampling time in early 2016 to the present, it would underline the influx of multidrug resistant K. pneumonia supporting the presented data.

We don't know the exact numbers here, so we are reluctant to elborate further. The hospital technicians in Kakamega estimate that the number of patients infected with MDR Klebsiella is increasing, but those estimates are not based on recorded numbers. One of the hazards of with working in a resource-limited environment such as rural Kenya is that unless specifically directed to keep them, potentially informative longitudinal epidemiological records are not routinely maintained. This is one of the many reasons why we know so little about how AMR is disseminated in Africa…….

2. Presentation of results

The amendments to the result section are satisfactory.

*Figure 1 is a good representation of the results. However, the genes in the black box are confusing. They are described as: "Sequence variants of marker genes traditionally used for MLST are depicted by framed boxes of different colour."

The genes seem to be MLST marker genes and virulence genes; aerobactin is already included in the virulence score. It is confusing that the ST and capsule types have overlapping colours which makes the Figure difficult to read.

FIGURE AMENDED as requested

*Ln 188 Hospital- and community-acquired Klebsiella pathogens were isolated

Amended

3. How the style and organization of the paper communicates and represents key findings

The amendments to the discussion section are satisfactory.

*Ln 327 currently change to at the time of publication

How much was it during sampling time? See comment above.

DONE

VERSION 2

Editor recommendation and comments

https://doi.org/10.1099/acmi.0.000667.v2.5
© 2023 Munnoch J. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License.

John Munnoch; University of Strathclyde, SIPBS, UNITED KINGDOM, Glasgow

Date report received: 01 December 2023
Recommendation: Minor Amendment

Comments: Thank you for your revisions, reviewers were happy with your responses, there are only now minor changes required (for example some software versions). No additional data processing is required, however where possible please include the information requested.

Reviewer 2 recommendation and comments

https://doi.org/10.1099/acmi.0.000667.v2.4
© 2023 Anonymous. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License.

Anonymous.

Date report received: 30 November 2023
Recommendation: Minor Amendment

Comments: 1. Methodological rigour, reproducibility and availability of underlying data The Methods section is now written clear and the reanalysis using Kleborate suggested by reviewer one adds valuable information to the manuscript. * The information of how many isolates were tested for AMR in total and what is the percentage of multi-drug resistant isolates is missing in the method section. This is only stated in the Results section. The information is on Ln122. Given that the % resistant isolates is a result, rather than a method, we would prefer to leave this where it currently is (Ln200). It would be beneficial to know how many K. pneumonia strains were isolated during the sampling time and how many of those are multi drug resistant. The authors stated that 5-10 patients a day present with K. pneumonia currently. Is that 2023? Is there a change in the presents of multidrug resistance K. pneumonia? If there is an increase from the sampling time in early 2016 to the present, it would underline the influx of multidrug resistant K. pneumonia supporting the presented data. 2. Presentation of results The amendments to the result section are satisfactory. *Figure 1 is a good representation of the results. However, the genes in the black box are confusing. They are described as: "Sequence variants of marker genes traditionally used for MLST are depicted by framed boxes of different colour." The genes seem to be MLST marker genes and virulence genes; aerobactin is already included in the virulence score. It is confusing that the ST and capsule types have overlapping colours which makes the Figure difficult to read. *Ln 188 Hospital- and community-acquired Klebsiella pathogens were isolated 3. How the style and organization of the paper communicates and represents key findings The amendments to the discussion section are satisfactory. *Ln 327 currently change to at the time of publication How much was it during sampling time? See comment above.

Please rate the manuscript for methodological rigour
Good

Please rate the quality of the presentation and structure of the manuscript
Satisfactory

To what extent are the conclusions supported by the data?
Partially support

Do you have any concerns of possible image manipulation, plagiarism or any other unethical practices?
No

Is there a potential financial or other conflict of interest between yourself and the author(s)?
No

If this manuscript involves human and/or animal work, have the subjects been treated in an ethical manner and the authors complied with the appropriate guidelines?
Yes

Reviewer 1 recommendation and comments

https://doi.org/10.1099/acmi.0.000667.v2.3
© 2023 Connor C. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License.

Christopher Connor; The Peter Doherty Institute for Infection and Immunity, The University of Melbourne, 792 Elizabeth St, Melbourne, AUSTRALIA
https://orcid.org/0000-0002-1761-2412

Date report received: 13 November 2023
Recommendation: Accept

Comments: I am happy that the authors have sufficiently addressed all my comments from the previous round of review. Their work contributes to the surveillance of an important pathogen in an under-reported region. Software version numbers for CheckM2, Kleborate and Kaptive are missing and should be included.

Please rate the manuscript for methodological rigour
Very good

Please rate the quality of the presentation and structure of the manuscript
Good

To what extent are the conclusions supported by the data?
Strongly support

Do you have any concerns of possible image manipulation, plagiarism or any other unethical practices?
No

Is there a potential financial or other conflict of interest between yourself and the author(s)?
No

If this manuscript involves human and/or animal work, have the subjects been treated in an ethical manner and the authors complied with the appropriate guidelines?
Yes

SciScore report

https://doi.org/10.1099/acmi.0.000667.v2.1
© 2023 The Authors. This is an open-access article report distributed under the terms of the Creative Commons License.

iThenticate report

https://doi.org/10.1099/acmi.0.000667.v2.2
© 2023 The Authors. This is an open-access article report distributed under the terms of the Creative Commons License.

Author response to reviewers to Version 1

Dear Dr Munnoch,

Thank you for sending our submission "Whole genome sequencing and genotyping *Klebsiella pneumoniae*multi-drug resistant hospital isolates from Western Kenya"

by Dinda *et al*out to review, and my apologies that the revision has taken some time to turn around…. Prof Reva and myself have been travelling a lot over the late summer/early autumn period, hence the delay.

We thank the referees for their detailed, insightful, and above all, *extremely helpful*comments, and we have now carried out a thorough re-analysis of the data using the approaches they suggest. Although the main conclusions have not changed as a consequence of this, the rigor of the analysis is undoubtedly better. Our point-by-point responses to the comments are given below (our responses in **red**, to aid visualisation). Note that because we have extensively modified the text, we have not used "track changes" function since it looked messy. Instead, we highlight the specific lines and changes made in the text below. We hope that this is OK.

Reviewer 1 Comments to Author:

Summary:

The article by Dinda et al. presents data on a collection of K. pneumoniae isolates gathered from a Kenyan hospital over a 6-month period circa. 2016. They identify isolates that display alarming levels of antibiotic resistance and proceed to whole genome sequence and genomically characterise them. They find a number of resistance genes present in the isolates and locate them on genomic islands and plasmids. Their primary finding is that the ST147 lineage has become a dominant strain in Kenyan hospitals. This article provides much needed phenotypic and genotypic observations on a major MDR pathogen from an under-surveyed region.

1. Methodological rigour, reproducibility and availability of underlying data

Several analyses which are considered standard by the field for this pathogen are missing from this article. All these analyses can be performed using Kleborate (https://github.com/klebgenomics/Kleborate) which is the standard analysis pipeline for this pathogen. This tool is even implemented in Pathogen Watch (https://pathogen.watch/). Analyses that are specifically missing: No genomic confirmation of species. The isolates in this study were identified as K. pneumoniae by culture and biochemical assay only. It has been previously published that biochemical and even MALDI-TOF analysis is not capable of accurately discriminating between Klebsiella species (see: 10.1128/mSphereDirect.00290-17). I'd like to see the genomic data used to confirm species with Kleborate, or even Kraken2 + GTDB (https://github.com/DerrickWood/kraken2& https://gtdb.ecogenomic.org/).

We thank the reviewer for recommending Kleborate. This is a very useful tool, indeed. The analysis was done and the following text was added:

Ln. 206-209: "All the isolates were unambiguously identified as *Klebsiella pneumoniae*strains by analysing their genome sequences with Kleborate software tool and by BLASTP mapping predicted GyrA sequences against the nr NCBI database."

No data were presented on virulence factors. The authors state that isolates were characterised using the Virulence Factor Database yet no results for virulence factors are presented. VFDB is a multi-species virulence factor database and not optimised for Klebsiella, again Kleborate would be more appropriate.

Virulence and antibiotic resistance factors are now presented in section 3.4. This section includes the predictions by Kleborate:

Ln. 267-275: "An overall prediction of the virulence and antibiotic resistance status of the isolates was performed by Kleborate software tool (Fig. 1 and Suppl. Table S2). Two strains, CK3 and CK8 isolated from the hospital and the community, were predicted as potentially dengerouse isolates due to possessing both aerobactin and yersiniabactin iron acquisition systems contributing significantly to virulence of Klebsiella and Escherichia [27]-[29]. Other 9 strains also possessed yersiniabactin, and 4 isolates had no these virulence factors. All the strains except for two (CK 1 and 2) possessed extended-spectrum beta-lactamases (ESBL, CTX-M-15) [30] including carbapenemases predicted by Kleborate in four strains (OXA-181 in strain K10, K11 and CK14; and NDM-1 in CK5), which render resistance to carbapenem antibiotics [31]."

The isolates have not been capsule typed. The capsule is important for both virulence and AMR in Klebsiella. Moreover, it is particularly relevant for ST147 as previous studies have identified specific K loci associated with this lineage. Again, Kleborate can perform capsule typing.

Capsule types identified by Kleborate/Kaptive were presented along with the MLST typing:

Ln. 211-217: ". Nine out of the 15 isolates were assigned to ST147 (including a subtype ST147-1LV, where 1LV indicates an additional SNP from previously published ST). The ST147 isolates were split in the phylogenetic tree into several clusters characterized by two different polysaccharide capsule types KL64 and KL10. Two isolates CK3 and CK8 belonged to ST231 and ST231-1LV with the same capsule type KL51. The remaining four minor isolates were assigned to ST11, ST14, ST1634 and ST1801 characterized by three different capsule types (Fig. 1)."

There is a mismatch between results and methods, were the genomes annotated with Prokka as per the methods section or with RAST as per the results section? If RAST was used it is not cited and the version number is absent.

An explanation was added to the Methods:

Ln. 143-147: "Preliminary genome annotation was done using the RAST Annotation Server (https://rast.nmpdr.org/) [5] with automatic fix error function that checks and fixes possible sequencing and genome assembly errors. The NCBI Prokaryotic Genome Annotation Pipeline (PGAP v.6.6) [9] was later used to obtain a more refined genome annotation."

It is not clear to the reader how the authors distinguish between community and hospital acquired infections.

This is now clarified [Ln 109-111] "Infections were designated as hospital- or community-acquired to distinguish between different possible sources of infection. The nosocomial infections were defined as those acquired by patients after hospitalization, manifesting 48 h after admission."

I commend the authors for making their genome assemblies publicly available with the publication however I would also like the raw FASTQ files to be made available as well. The genome assembly records can be linked to SRA records through the Biosample ID's. The FASTQ files are considered the raw data and their availability contributes to open access, reproducible science. The genome assemblies are not considered as raw data. NCBI Biosamples also allow submission of antibiogram data such as zones of inhibition from disk diffusion assays, making this data available (if possible) would be highly beneficial to the research community.

The raw reads have now been uploaded (see below) and the BioSamnple IDs are listed in Table 1.

BioProject BioSample Strain SRA Accession

PRJNA791700 SAMN24338476 CK1 SRR26542240

PRJNA896788 SAMN31566495 CK2 SRR26542242

PRJNA896856 SAMN31566649 CK3 SRR26544990

PRJNA791929 SAMN24369220 CK4 SRR26545085

PRJNA791933 SAMN24369239 CK5 SRR26553297

PRJNA791937 SAMN24369281 CK6 SRR26545104

PRJNA791940 SAMN24369410 CK7 SRR26545418

PRJNA791951 SAMN24370162 CK8 SRR26545506

PRJNA791957 SAMN24371033 CK9 SRR26545513

PRJNA791958 SAMN24371047 CK10 SRR26547826

PRJNA792094 SAMN24377653 K11 SRR26546318

PRJNA792103 SAMN24377668 K12 SRR26548528

PRJNA792129 SAMN24386524 K13 SRR26548540

PRJNA792139 SAMN24386561 K14 SRR26549125

PRJNA792142 SAMN24386923 K15 SRR26548991

There was no mention of quality control on the genomic data. Illumina reads were trimmed with Trimmomatic but there was no assessment of levels of contamination though Kraken2 or quantification of sequencing coverage; were the isolates sequenced at 30x coverage or 100x coverage? There was also no assessment of resulting genome assembly metrics (number of contigs, genome size, N50 etc.), these metrics could be calculated with CheckM (https://github.com/Ecogenomics/CheckM) or Quast (https://github.com/ablab/quast).

CheckM2 metrics have now been added to Table 1. Also see genome assembly metrics produced by Kleborate in Table S2. Depth of sequencing is now stated on line 135.

The authors use the tool SeqWord Genomic Island Sniffer to identify Genomic Islands, I am unfamiliar with this tool. The cited article is not readily available (no results from a Google or Pubmed search), nor could I locate a code repository for this tool (available hyperlinks returned errors). The paper would benefit from a brief overview of the tool's methodology in predicting horizontally transferred islands, as well as providing a link to the tool. Were there any additional analyses performed to confirm predictions such as identifying transposons at genomic island borders? Or prediction of phage?

The web address for SeqWord has now been added and we have confirmed that the link is correct: http://seqword.bi.up.ac.za/sniffer/index.html

We have also now added a line stating how SeqWord operates (line 182: "SeqWord identifies horizontally-acquired genomic elements by analyzing patterns of oligonucleotide signatures.")

Version numbers and parameters for software and databases should be listed, as well as cited. Some examples: the Mauve version number is missing and is not cited. RAST (if used, see comment above) is also missing a version number and citation. Resistance Gene Identifier (RGI) version number. CARD database version or date of access. MUSCLE parameters used and version number. Gblocks version, etc.

We have now indicated the software version used throughout. However, we note that web-based services such as RAST are not always associated with correct current version numbers.

The phylogenetic method used is unusual for such a small set of closely related isolates. The authors opt to produce a peptide alignment and build a Neighbour Joining tree from this, while this approach is valid a more typical approach used in the field would be a Maximum Likelihood tree from a nucleotide core genome alignment. This could be achieved by read mapping to a reference genome to produce a core genome alignment with Snippy (https://github.com/tseemann/snippy). A phylogenetic tree could then be built using IQTree or RAxML (IQTree: http://www.iqtree.org/RAxML: https://github.com/amkozlov/raxml-ng). Another option is the tool Parsnp which performs both alignment and tree construction (https://github.com/marbl/parsnp). The peptide alignment of orthologs is perhaps better suited to a more diverse dataset spanning multiple species / genera.

As the referee notes, our approach is a valid one. We used this approach (identification of orthologous protein coding sequences followed by concatenation of alignments) in a previous study (https://doi.org/10.1016/j.meegid.2021.104784) and we want to be consistent. Maximum Likelihood is not applicable for this approach as the alignment is too long (805,265 amino acid residues). Moreover, and given our goal was just to group the strains into clusters, we note that NJ is an acceptable phylogenetic clustering algorithm. As shown in Fig. 1, the resulting tree matches very well with the genotyping data produced by Kleborate.

2. Presentation of results

The isolates assayed display resistance to the carbapenem antibiotics Imipenem and Meropenem yet a carbapenemase was not found? Could the authors comment on this observation? Was carbapenem resistance borderline? What mechanism do they think is driving resistance? See: 10.1128/JCM.00651-08. In Supplementary Table 1 the carbapenemases NDM-6 and OXA-232 are listed. Why were these results not discussed in the main body? The discovery of carbapenemase genes on plasmids is an important observation.

These are good points, and following our re-analysis of the data using Kleborate, we have now included the following text:

Ln. 269-278: "An overall prediction of the virulence and antibiotic resistance status of the isolates was performed using Kleborate (Fig. 1 and Table S2). Two strains, CK3 and CK8 (both ST231) isolated from the hospital and from the community (respectively), were predicted to encode both the aerobactin and yersiniabactin iron acquisition systems. These siderophore systems contribute

significantly to the virulence of *Klebsiella*and *Escherichia*sp. **Error! Reference source not found.**-**Error! Reference source not found.**. Another nine strains possessed the yersiniabactin system alone, and four isolates encoded neither siderophore system. All strains except for two (CK1 and CK2) encoded extended-spectrum b-lactamases (ESBL, CTX-M-15) **Error! Reference source not found.**and Kleborate predicted carbapenemases in four strains (OXA-181 in strain K10, K11 and CK14; and NDM-1 in CK5), rendering resistance to carbapenem antibiotics **Error! Reference source not found.**.."

The software used to produce the figures is not clearly stated. Was iTOL (https://itol.embl.de/) used to create the trees? Or were they produced in ggtree (https://github.com/YuLab-SMU/ggtree)? Or other software?

We have now included these data in the Methods:

Ln. 165-166: "…further phylogenetic inferences using the Neighbour-Joining algorithm implemented MEGA-X [14]."

Ln. 178-180: "The Kleborate and Kaptive predictions were combined with the phylogenetic tree and visualized using the Microreact Web-service (https://microreact.org/)."

It is also not clear to the reader what data the dendrogram in Figure 2 is based on. What distance metric was used? Blast identity? Coverage? Average nucleotide identity? What does the scale bar represent? It would also be beneficial to add virulence gene results to this figure.

We have now modified the Methods to clarify how the dendrogram in Fig 2 was generated. A detailed account of Mauve is provided, as noted, in Ref 15. The revised text now states;

Ln. 166-171: "Alignment of plasmid sequences was performed by Mauve 20150226 [15]. Here, the sequence similarity is the ratio of base pairs shared between two genomes to their average genome length. This similarity estimate is then converted to a distance value for the NJ distance matrix. The progressiveMauve algorithm produces a dendrogram of sequence similarity relations stored in an alignment.guide_tree file. The dendrogram was visualized by MEGA-X."

Supplementary Figure 1 is difficult to interpret. What do the colour blocks represent? What are the lines connecting the plasmid contigs? I'd suggest the plasmid contigs identified are grouped based on some characteristic such as size or gene content (virulence / AMR) and then compared as separate panels.

We have now changed the legend to the figure to clarify these points:

Ln 374-379: **Figure S1**. Alignment of sequences of the plasmids from different strains of *K. pneumoniae*produced by Mauve 20150226. Each sequence is represented by a row with a central line and sequence length values. Blocks filled with the same colour and connected by thin vertical lines represent homologous sequences found in different plasmids. Blocks above the central line are located on the forward strand and those below the line are located on the reverse complement strand of the plasmids.

3. How the style and organization of the paper communicates and represents key findings

Overall, the paper is well written and has a clear flow of reasoning. However, a key discovery of carbapenemase genes located on plasmids is not discussed in the main body of the paper.

Amended – see above.

Some results (strain details, gene detection) would be better suited to presentation in tables and not listed extensively in the paper (Section 3.4).

Much of this has now been moved to Table S1, as requested.

The conclusion that ST147 is the dominant lineage from such a small sample collection over a limited time period from a single hospital is premature. Especially given that the authors only sequenced 15 out of the 27 isolates collected, focussing on the most resistant ones. I'd recommend the authors amend their conclusion to reflect this limitation, perhaps clarifying that ST147 is the dominant lineage in highly resistant K. pneumoniae.

Good point. We have now amended the main conclusion (alluded to in the Discussion and the Abstract) to state that ST147 is the dominant lineage in highly resistant *K. pneumoniae*.

4. Literature analysis or discussion

The authors aptly describe the threat posed by multi-drug resistant K. pneumoniae and the lack of surveillance data from certain regions. The authors discuss their findings well, particularly with results from recent studies of the same region. They postulate that human migration or pathogenesis may drive the success of ST147.

No response required.

5. Any other relevant comments

Typos and lack of clarity:

I would also recommend using the acronym Antibiotic / Antimicrobial Resistance Genes (ARG's) instead of Drug Resistance Genes (DRG) to be more consistent with the literature.

- Done

Line 148 and 198: correct GRI-CARD to RGI-CARD. Clearer to specify that Resistance Gene Identifier (RGI) tool was used with the CARD database. Tool and database versions are also missing.

- Done

Line 275: "the infection is either patchy and differs from clinic to clinic" - unclear what this statement means. Do the authors mean that there are geographic differences in the presence of ST147?

- Yes, this statement was confusing and did not add anything. It has now been removed.

Line 40, 154 and 262: Correct pneumonia to pneumonia

- Done

Line 55: Why is there a question mark after SPAdes version number?

- Removed

Line 264: correct multigrug to multidrug

- Done

Line 274: correct discrepensy to discrepancy

- Done

Line 276: correct resent to recent

- Fixed

Line 284: correct plismids to plasmids

- Done

Line 288: correct turkey to Turkey / Türkiye

- This sentence was deleted in the revised version, so the comment no longer applies.

Line 427: correct resustance to resistance

- Done

The comments below should be considered as optional. I believe addressing them could elevate the article.

The authors mention an increase in MDR K. pneumoniae infections triggered a targeted investigation, more information on typical case numbers seen at the hospital would be beneficial. The authors state that they performed AST on 27 isolates, is 27 K. pneumoniae in a 6-month period typical? If not, what are the typical numbers?

KCGTRH currently receives 5-10 patients per day with *K. pneumoniae*-associated infections.This is now stated in the Discussion (line 327). Of course, not all of these exhibit multidrug resistance….

A more systematic comparison of phenotypic and genotypic resistance profiles would be beneficial. Specifically, can all the phenotypic resistances be explained by genomic factors? Do all the pan-resistant isolates contain genes conferring resistance to each antibiotic tested? Are there any isolates with unexplained phenotypic resistance?

Many factors contribute towards the "resistome" for any given antibiotic. Moreover, it should be noted that finding an ARG in the genome does not necessarily account for the levelof antibiotic resistance – that requires detailed analyses of expression levels and copy number levels beyond the scope of the current work.

Further characterisation of the plasmid contigs identified. Specifically, incompatibility group typing, replicon identification and conjugation machinery. The tool MOB-typer would be useful for this (https://github.com/phac-nml/mob-suite). Moreover, K. pneumoniae virulence and AMR determinants are known to circulate on plasmids, the convergence of these genes onto the same plasmid is of great concern and subject to surveillance. More detailed plasmid characterisation with respect to virulence and AMR gene content would be beneficial; do the plasmids harbouring carbapenemase genes also possess virulence factors or conjugation machinery? Did the plasmid contigs identified show any similarity to previously published plasmids? A blast database search would identify similar sequences.

This was a great suggestion. The plasmids were re-analysed by Mob-Suite and we revised the text accordingly (Ln. 290-295) to state: "A search for similar plasmids in the NCBI was perfomed using Mob-Suite software **Error! Reference source not found.**. All the plasmids were predicted by this tool as conjugative broad-range plasmids belonging to mate-pair formation (MPF) types I and F. MPF classification of the plasmids is in agreement with their clustering in the dendrogram (Fig. 2). The full list of determined Mob-Suite metrics is in Table S3"

Reviewer 2 Comments to Author:

1. Methodological rigour, reproducibility and availability of underlying data

* The information of how many isolates were tested for AMR in total and what is the percentage of multi-drug resistant isolates is missing in the method section. This is only stated in the Results section.

The information is on Ln122. Given that the % resistant isolates is a result, rather than a method, we would prefer to leave this where it currently is (Ln200).

* The isolates were obtained by standard techniques and ethics approval was appropriate.

* Genome sequencing and annotation was performed using standard methods. However, the authors state in the data summary that the DNA reads were quality controlled but do not provide the method.

We have now added this to the text;

Ln. 137-138: "The reads were trimmed using Trimmomatic v0.30 with a sliding window quality cut-off of Q15."

Ln. 153-154: "Quality of the resilting consensus sequences was controlled by the program CheckM2 [6]."

Ln. 211: "Genome completeness and purity were confirmed by CheckM2 analysis (Table 1)."

* The genome data is deposited on the NCBI server. The authors state in the method section that the genomes were annotated using Prokka v1.14.5. Contrary, the annotation details on the NCBI website state that the genome was annotated using NCBI Prokaryotic Genome Annotation Pipeline (PGAP). Additionally, in the results section, line 183, RAST genome annotation robot was used as basis for the phylogenetic tree. Hence, it is not clear which annotation methods was used to analyse the data.

See response to referee 1.

* The authors performed a search for virulence genes stated in the methodology line 148 but do not show any results regarding virulence genes.

This is a good point, and one well made. We have now included a segment on virulence genes in the Results;

Ln. 272-287: "An overall prediction of the virulence and antibiotic resistance status of the isolates was performed using Kleborate (Fig. 1 and Table S2). Two strains, CK3 and CK8 (both ST231) isolated from the hospital and from the community (respectively), were predicted to encode both the aerobactin and yersiniabactin iron acquisition systems. These siderophore systems contribute significantly to the virulence of *Klebsiella*and *Escherichia*sp. **Error! Reference source not found.**-**Error! Reference source not found.**. Another nine strains possessed the yersiniabactin system alone, and four of the isolates encoded neither siderophore system. All strains except for two (CK1 and CK2) encoded extended-spectrum b-lactamases (ESBL, CTX-M-15) **Error! Reference source not found.**and Kleborate predicted carbapenemases in four strains (OXA-181 in strains K10, K11 and CK14; and NDM-1 in CK5) rendering resistance to carbapenem antibiotics **Error! Reference source not found.**. The streptomycin resistance genes, *strAB*, and the sulfonamide resistant dihydropteroate synthase, *sul2*, were found in CK5, K10, K11, K12, K13, and K15. Strain K14 also possessed an aminoglycoside-modifying enzyme, encoded by *sat*-2. Mutations in *gyrA*and *parC*associated with fluoroquinolone resistance **Error! Reference source not found.**were identified in all strains except for CK1 and CK2. Finally, CK3, CK4, CK8, K10, K11, K12, K13, K14 and K15 all carried mutations in the outer membrane porins, *ompK35*and *ompK36*, associated with doripenem-colistin resistance **Error! Reference source not found.**."

* The table with NCBI accession number is more appropriate in the supplementary data.

We prefer to keep the table in the main text now that more information from the CheckM2 assembly quality control has been added.

2. Presentation of results

* Figure 1 shows the phylogenetic tree of sequenced isolates. It is a good representation of the relationship of the isolates. The authors write that five isolates were sensitive to one respective tested antibiotic. It would be interesting to add this data to the phylogenetic tree of the isolates and further which drug related genes (DRG) relating to the sensitive antibiotic is missing in these isolates either in the genome or plasmid.

Done.

* Figure 2 was created using Mauve alignment. The figure shows a root, but it is not clear which sequence that is.

The tree legend now states: "The tree was rooted at the average branch location not representing the common ancestry."

* The authors make the intuitive assumption that the larger plasmids in K10, K12, K13 and K15 evolved from the smaller ancestral plasmids found in CK2 and CK4. Are there any transposable elements on the plasmids to support this hypothesis? It is also possible that the ancestral plasmid was larger in size and lost DRG in CK2 and CK4.

* Similar conclusions were made for the phylogenetic group containing plasmids CK9, CK7, CK6 and CK1. The larger plasmids and smaller plasmids share a common ancestor. It is difficult to say if the larger plasmids evolved from the smaller ones with the small sample size. The common ancestor plasmid could have lost some DRG.

There are transposable elements, although these are not proof of the suggested direction of evolution. This is now discussed in Ln. 309-315: "Assuming that the evolution of virulence plasmids in the antibiotic era is towards rendering multidrug resistance, these plasmids may have evolved from the smaller, ancestral plasmids represented by those in isolates CK2 (ST1634) and CK4 (ST147) through the acquisition of additional ARGs. The ST147 isolates CK6, CK7 and CK9 contain another type of plasmid (size range 301-335 kb). These plasmids may also have evolved from smaller plasmids, although here, the ancestor appears to be similar to the plasmid in isolate CK1 (ST1801)."

* Figure 2 could be improved by adding the presents or absence of DGR in a heatmap. This heatmap could also include virulence genes.

This information is provided in Supplementary Table S2. Additionally, we used the Mob-Suite to characterize the plasmids in more detail (as per the suggestion from referee 1). See Ln. 289-294: "A search for similar plasmids in the NCBI was perfomed using

Mob-Suite software **Error! Reference source not found.**. All the plasmids were predicted by this tool as conjugative broad-range plasmids belonging to mate-pair formation (MPF) types I and F. MPF classification of the plasmids is in agreement with their clustering in the dendrogram (Fig. 2). The full list of determined Mob-Suite metrics is in Table S3."

3. How the style and organization of the paper communicates and represents key findings

* The authors use standard techniques to demonstrate the importance of WGS as a surveillance tool for AMR in under surveyed regions. They clearly demonstrate the presence of multi-drug resistant K. pneumonia strains. However, the evolutionary pathway of the acquisition of DGR needs to be investigated further. This could be achieved by including additional genome and/or plasmid sequences in the phylogenetic trees.

See comments above.

* The authors state that the evolutionary success of the investigated isolates is because of the acquisition of DGR genes and virulence genes. They present data of DGR confirming this but miss to show how virulence genes support this hypothesis.

As outlined above, we have now addressed this comment by providing new data following the reanalysis done using Kleborate.

4. Literature analysis or discussion

* Line 261: Do the authors mean references [23] and [24] instead of [25]?

Amended.

* The distribution of ST needs to be discussed more in a geographical and temporal context. Reference [25] was published in 2022 but sampling occurred from May 2015 till March 2020.

Good point. We have now re-worked the Discussion (paragraph 2) to take into account geographical and temporal issues.

* ST147 and ST14 have been reported to spread DRG. This could be included to discuss the evolutionary success of the sequenced isolates.

Again, we have re-worked the Discussion (paragraph 2) to address the formidible capacity of ST147 to disseminate AMR determinants.

VERSION 1

Editor recommendation and comments

https://doi.org/10.1099/acmi.0.000667.v1.5
© 2023 Munnoch J. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License.

John Munnoch; University of Strathclyde, SIPBS, UNITED KINGDOM, Glasgow

Date report received: 21 August 2023
Recommendation: Major Revision

Comments: This study would be a valuable contribution to the existing literature. This is a study that would be of interest to the field and community. The reviewers have highlighted major concerns with the work presented. Please ensure that you address their comments.

Reviewer 2 recommendation and comments

https://doi.org/10.1099/acmi.0.000667.v1.4
© 2023 Som N. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License.

Nicolle Som; UNITED KINGDOM
https://orcid.org/0009-0000-6191-6846

Date report received: 13 August 2023
Recommendation: Minor Amendment

Comments: 1. Methodological rigour, reproducibility and availability of underlying data * The information of how many isolates were tested for AMR in total and what is the percentage of multi-drug resistant isolates is missing in the method section. This is only stated in the Results section. * The isolates were obtained by standard techniques and ethics approval was appropriate. * Genome sequencing and annotation was performed using standard methods. However, the authors state in the data summary that the DNA reads were quality controlled but do not provide the method. * The genome data is deposited on the NCBI server. The authors state in the method section that the genomes were annotated using Prokka v1.14.5. Contrary, the annotation details on the NCBI website state that the genome was annotated using NCBI Prokaryotic Genome Annotation Pipeline (PGAP). Additionally, in the results section, line 183, RAST genome annotation robot was used as basis for the phylogenetic tree. Hence, it is not clear which annotation methods was used to analyse the data. * The authors performed a search for virulence genes stated in the methodology line 148 but do not show any results regarding virulence genes. *The table with NCBI accession number is more appropriate in the supplementary data. 2. Presentation of results * Figure 1 shows the phylogenetic tree of sequenced isolates. It is a good representation of the relationship of the isolates. The authors write that five isolates were sensitive to one respective tested antibiotic. It would be interesting to add this data to the phylogenetic tree of the isolates and further which drug related genes (DRG) relating to the sensitive antibiotic is missing in these isolates either in the genome or plasmid. * Figure 2 was created using Mauve alignment. The figure shows a root, but it is not clear which sequence that is. * The authors make the intuitive assumption that the larger plasmids in K10, K12, K13 and K15 evolved from the smaller ancestral plasmids found in CK2 and CK4. Are there any transposable elements on the plasmids to support this hypothesis? It is also possible that the ancestral plasmid was larger in size and lost DRG in CK2 and CK4. * Similar conclusions were made for the phylogenetic group containing plasmids CK9, CK7, CK6 and CK1. The larger plasmids and smaller plasmids share a common ancestor. It is difficult to say if the larger plasmids evolved from the smaller ones with the small sample size. The common ancestor plasmid could have lost some DRG. * Figure 2 could be improved by adding the presents or absence of DGR in a heatmap. This heatmap could also include virulence genes. 3. How the style and organization of the paper communicates and represents key findings * The authors use standard techniques to demonstrate the importance of WGS as a surveillance tool for AMR in under surveyed regions. They clearly demonstrate the presence of multi-drug resistant K. pneumonia strains. However, the evolutionary pathway of the acquisition of DGR needs to be investigated further. This could be achieved by including additional genome and/or plasmid sequences in the phylogenetic trees. * The authors state that the evolutionary success of the investigated isolates is because of the acquisition of DGR genes and virulence genes. They present data of DGR confirming this but miss to show how virulence genes support this hypothesis. 4. Literature analysis or discussion * Line 261: Do the authors mean references [23] and [24] instead of [25]? * The distribution of ST needs to be discussed more in a geographical and temporal context. Reference [25] was published in 2022 but sampling occurred from May 2015 till March 2020. * ST147 and ST14 have been reported to spread DRG. This could be included to discuss the evolutionary success of the sequenced isolates.

Please rate the manuscript for methodological rigour
Good

Please rate the quality of the presentation and structure of the manuscript
Satisfactory

To what extent are the conclusions supported by the data?
Partially support

Do you have any concerns of possible image manipulation, plagiarism or any other unethical practices?
No

Is there a potential financial or other conflict of interest between yourself and the author(s)?
No

If this manuscript involves human and/or animal work, have the subjects been treated in an ethical manner and the authors complied with the appropriate guidelines?
Yes

Reviewer 1 recommendation and comments

https://doi.org/10.1099/acmi.0.000667.v1.3
© 2023 Connor C. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License.

Christopher Connor; The Peter Doherty Institute for Infection and Immunity, The University of Melbourne, 792 Elizabeth St, Melbourne, AUSTRALIA
https://orcid.org/0000-0002-1761-2412

Date report received: 20 July 2023
Recommendation: Major Revision

Comments: Summary: The article by Dinda et al. presents data on a collection of K. pneumoniae isolates gathered from a Kenyan hospital over a 6-month period circa. 2016. They identify isolates that display alarming levels of antibiotic resistance and proceed to whole genome sequence and genomically characterise them. They find a number of resistance genes present in the isolates and locate them on genomic islands and plasmids. Their primary finding is that the ST147 lineage has become a dominant strain in Kenyan hospitals. This article provides much needed phenotypic and genotypic observations on a major MDR pathogen from an under-surveyed region. 1. Methodological rigour, reproducibility and availability of underlying data Several analyses which are considered standard by the field for this pathogen are missing from this article. All these analyses can be performed using Kleborate (https://github.com/klebgenomics/Kleborate) which is the standard analysis pipeline for this pathogen. This tool is even implemented in Pathogen Watch (https://pathogen.watch/). Analyses that are specifically missing: No genomic confirmation of species. The isolates in this study were identified as K. pneumoniae by culture and biochemical assay only. It has been previously published that biochemical and even MALDI-TOF analysis is not capable of accurately discriminating between Klebsiella species (see: 10.1128/mSphereDirect.00290-17). I'd like to see the genomic data used to confirm species with Kleborate, or even Kraken2 + GTDB (https://github.com/DerrickWood/kraken2 & https://gtdb.ecogenomic.org/). No data were presented on virulence factors. The authors state that isolates were characterised using the Virulence Factor Database yet no results for virulence factors are presented. VFDB is a multi-species virulence factor database and not optimised for Klebsiella, again Kleborate would be more appropriate. The isolates have not been capsule typed. The capsule is important for both virulence and AMR in Klebsiella. Moreover, it is particularly relevant for ST147 as previous studies have identified specific K loci associated with this lineage. Again, Kleborate can perform capsule typing. There is a mismatch between results and methods, were the genomes annotated with Prokka as per the methods section or with RAST as per the results section? If RAST was used it is not cited and the version number is absent. It is not clear to the reader how the authors distinguish between community and hospital acquired infections. I commend the authors for making their genome assemblies publicly available with the publication however I would also like the raw FASTQ files to be made available as well. The genome assembly records can be linked to SRA records through the Biosample ID's. The FASTQ files are considered the raw data and their availability contributes to open access, reproducible science. The genome assemblies are not considered as raw data. NCBI Biosamples also allow submission of antibiogram data such as zones of inhibition from disk diffusion assays, making this data available (if possible) would be highly beneficial to the research community. There was no mention of quality control on the genomic data. Illumina reads were trimmed with Trimmomatic but there was no assessment of levels of contamination though Kraken2 or quantification of sequencing coverage; were the isolates sequenced at 30x coverage or 100x coverage? There was also no assessment of resulting genome assembly metrics (number of contigs, genome size, N50 etc.), these metrics could be calculated with CheckM (https://github.com/Ecogenomics/CheckM) or Quast (https://github.com/ablab/quast). The authors use the tool SeqWord Genomic Island Sniffer to identify Genomic Islands, I am unfamiliar with this tool. The cited article is not readily available (no results from a Google or Pubmed search), nor could I locate a code repository for this tool (available hyperlinks returned errors). The paper would benefit from a brief overview of the tool's methodology in predicting horizontally transferred islands, as well as providing a link to the tool. Were there any additional analyses performed to confirm predictions such as identifying transposons at genomic island borders? Or prediction of phage? Version numbers and parameters for software and databases should be listed, as well as cited. Some examples: the Mauve version number is missing and is not cited. RAST (if used, see comment above) is also missing a version number and citation. Resistance Gene Identifier (RGI) version number. CARD database version or date of access. MUSCLE parameters used and version number. Gblocks version, etc. The phylogenetic method used is unusual for such a small set of closely related isolates. The authors opt to produce a peptide alignment and build a Neighbour Joining tree from this, while this approach is valid a more typical approach used in the field would be a Maximum Likelihood tree from a nucleotide core genome alignment. This could be achieved by read mapping to a reference genome to produce a core genome alignment with Snippy

(https://github.com/tseemann/snippy). A phylogenetic tree could then be built using IQTree or RAxML (IQTree: http://www.iqtree.org/ RAxML: https://github.com/amkozlov/raxml-ng). Another option is the tool Parsnp which performs both alignment and tree construction (https://github.com/marbl/parsnp). The peptide alignment of orthologs is perhaps better suited to a more diverse dataset spanning multiple species / genera. 2. Presentation of results The isolates assayed display resistance to the carbapenem antibiotics Imipenem and Meropenem yet a carbapenemase was not found? Could the authors comment on this observation? Was carbapenem resistance borderline? What mechanism do they think is driving resistance? See: 10.1128/JCM.00651-08. In Supplementary Table 1 the carbapenemases NDM-6 and OXA-232 are listed. Why were these results not discussed in the main body? The discovery of carbapenemase genes on plasmids is an important observation. The software used to produce the figures is not clearly stated. Was iTOL (https://itol.embl.de/) used to create the trees? Or were they produced in ggtree (https://github.com/YuLab-SMU/ggtree)? Or other software? It is also not clear to the reader what data the dendrogram in Figure 2 is based on. What distance metric was used? Blast identity? Coverage? Average nucleotide identity? What does the scale bar represent? It would also be beneficial to add virulence gene results to this figure. Supplementary Figure 1 is difficult to interpret. What do the colour blocks represent? What are the lines connecting the plasmid contigs? I'd suggest the plasmid contigs identified are grouped based on some characteristic such as size or gene content (virulence / AMR) and then compared as separate panels. 3. How the style and organization of the paper communicates and represents key findings Overall, the paper is well written and has a clear flow of reasoning. However, a key discovery of carbapenemase genes located on plasmids is not discussed in the main body of the paper. Some results (strain details, gene detection) would be better suited to presentation in tables and not listed extensively in the paper (Section 3.4). The conclusion that ST147 is the dominant lineage from such a small sample collection over a limited time period from a single hospital is premature. Especially given that the authors only sequenced 15 out of the 27 isolates collected, focussing on the most resistant ones. I'd recommend the authors amend their conclusion to reflect this limitation, perhaps clarifying that ST147 is the dominant lineage in highly resistant K. pneumoniae. 4. Literature analysis or discussion The authors aptly describe the threat posed by multi-drug resistant K. pneumoniae and the lack of surveillance data from certain regions. The authors discuss their findings well, particularly with results from recent studies of the same region. They postulate that human migration or pathogenesis may drive the success of ST147. 5. Any other relevant comments Typos and lack of clarity: I would also recommend using the acronym Antibiotic / Antimicrobial Resistance Genes (ARG's) instead of Drug Resistance Genes (DRG) to be more consistent with the literature. Line 148 and 198: correct GRI-CARD to RGI-CARD. Clearer to specify that Resistance Gene Identifier (RGI) tool was used with the CARD database. Tool and database versions are also missing. Line 275: "the infection is either patchy and differs from clinic to clinic" - unclear what this statement means. Do the authors mean that there are geographic differences in the presence of ST147? Line 40, 154 and 262: Correct pneumonia to pneumoniae Line 55: Why is there a question mark after SPAdes version number? Line 264: correct multigrug to multidrug Line 274: correct discrepensy to discrepancy Line 276: correct resent to recent Line 284: correct plismids to plasmids Line 288: correct turkey to Turkey / Türkiye Line 427: correct resustance to resistance The comments below should be considered as optional. I believe addressing them could elevate the article. The authors mention an increase in MDR K. pneumoniae infections triggered a targeted investigation, more information on typical case numbers seen at the hospital would be beneficial. The authors state that they performed AST on 27 isolates, is 27 K. pneumoniae in a 6-month period typical? If not, what are the typical numbers? A more systematic comparison of phenotypic and genotypic resistance profiles would be beneficial. Specifically, can all the phenotypic resistances be explained by genomic factors? Do all the pan-resistant isolates contain genes conferring resistance to each antibiotic tested? Are there any isolates with unexplained phenotypic resistance? Further characterisation of the plasmid contigs identified. Specifically, incompatibility group typing, replicon identification and conjugation machinery. The tool MOB-typer would be useful for this (https://github.com/phac-nml/mob-suite). Moreover, K. pneumoniae virulence and AMR determinants are known to circulate on plasmids, the convergence of these genes onto the same plasmid is of great concern and subject to surveillance. More detailed plasmid characterisation with respect to virulence and AMR gene content would be beneficial; do the plasmids harbouring carbapenemase genes also possess virulence factors or conjugation machinery? Did the plasmid contigs identified show any similarity to previously published plasmids? A blast database search would identify similar sequences.

Please rate the manuscript for methodological rigour
Poor

Please rate the quality of the presentation and structure of the manuscript
Good

To what extent are the conclusions supported by the data?
Partially support

Do you have any concerns of possible image manipulation, plagiarism or any other unethical practices?
No

Is there a potential financial or other conflict of interest between yourself and the author(s)?
No

If this manuscript involves human and/or animal work, have the subjects been treated in an ethical manner and the authors complied with the appropriate guidelines?
Yes

SciScore report

https://doi.org/10.1099/acmi.0.000667.v1.1
© 2023 The Authors. This is an open-access article report distributed under the terms of the Creative Commons License.

iThenticate report

https://doi.org/10.1099/acmi.0.000667.v1.2
© 2023 The Authors. This is an open-access article report distributed under the terms of the Creative Commons License.