Longitudinal analysis within one hospital in sub-Saharan Africa over 20 years reveals repeated replacements of dominant clones of *Klebsiella pneumoniae* and stresses the importance to include temporal patterns for vaccine design considerations

*Eva Heinz[1,2], Oliver Pearse[2,3], Allan Zuza[3], Sithembile Bilima[3], Chisomo Msefula[4], Patrick Musicha[2,3], Patriciah Siyabu[5], Edith Tewesa[5], Fabrice E Graf[2], Rebecca Lester[3,6], Samantha Lissauer[3,7], Jennifer Cornick[3,7], Joseph M Lewis[2,3,7], Kondwani Kawaza[3,4], Nicholas R Thomson[8,9], *Nicholas A Feasey[2,3,10]

1. Department of Vector Biology, Liverpool School of Tropical Medicine, Liverpool, UK
2. Department of Clinical Sciences, Liverpool School of Tropical Medicine, Liverpool, UK
3. Malawi Liverpool Wellcome Programme, Kamuzu University of Health Sciences, Blantyre, Malawi
4. Kamuzu University of Health Sciences, Blantyre, Malawi
5. Queen Elizabeth Central Hospital, Blantyre, Malawi
6. Division of Infection & Immunity, University College London, UK
7. Department of Clinical Infection, Microbiology and Immunology, University of Liverpool, Liverpool, UK
8. Parasites and Microbes Program, Wellcome Sanger Institute, Hinxton, UK
9. London School of Hygiene and Tropical Medicine, London, UK
10. School of Medicine, St Andrews University, UK

*Contact: eva.heinz@lstmed.ac.uk, Nicholas.feasey@lstmed.ac.uk

Keywords: AMR, Malawi, genome surveillance, healthcare-associated, hospital-acquired, nosocomial infections, surface antigens, invasive, neonatal infection, sepsis, capsular polysaccharide, opportunistic pathogen, ESKAPE

Abstract

Infections caused by multidrug-resistant gram-negative bacteria present a severe threat to global public health. The WHO defines drug-resistant *Klebsiella pneumoniae* as a priority pathogen for which alternative treatments are needed given the limited treatment options and the rapid acquisition of novel resistance mechanisms by this species. Longitudinal descriptions of genomic epidemiology of *Klebsiella pneumoniae* can inform management strategies but data from sub-Saharan Africa are lacking.

We present a longitudinal analysis of all invasive *K. pneumoniae* isolates from a single hospital in Blantyre, Malawi, southern Africa, from 1998-2020, combining clinical data with genome sequence analysis of the isolates. We show that after a dramatic increase in the number of infections from 2016 *K. pneumoniae* becomes hyperendemic, driven by an increase in neonatal infections. Genomic data show repeated waves of clonal expansion of different, often ward-restricted, lineages, suggestive of hospital associated transmission. We describe temporal trends in resistance and surface antigens, of relevance for vaccine development.

Our data highlight a clear need for new interventions to prevent rather than treat *K. pneumoniae* infections in our setting. Whilst one option may be a vaccine, the majority of cases could be avoided by an increased focus on and investment in infection prevention and control measures, which would reduce all healthcare associated infections and not just one.

Introduction

The burden of deaths in children under the age of five, including those due to neonatal sepsis, is especially high in low- and middle-income countries (LMICs)(1); and one of the main pathogens causing these is *Klebsiella pneumoniae* (*Kpn*)(2). Until the last decade, *Kpn* was typically readily treatable with beta-lactam antimicrobials, in particular third generation cephalosporins (3GC). The effectiveness of 3GC is lost when these bacteria acquire extended-spectrum beta-lactamase enzymes (ESBLs). For *Kpn* there has been a drastic increase in 3GC-resistance by acquisition of ESBL enzyme-encoding plasmids; the proportion of *Kpn* hospital isolates that are ESBL-producing increased from 12% to >90% in parts of sub-Saharan Africa (sSA) in the last decade(3). The main remaining treatment option are carbapenems; these however are uncommonly available in sSA, rendering *Kpn* effectively untreatable in large parts of Africa. This is particularly problematic for neonatal sepsis caused by *Kpn*, and there is increasing evidence from large studies such as BARNARDS(2) and CHAMPS(4) that the burden of disease caused by *Kpn* falls most heavily on this patient group. Treatment options for neonates are even more limited than for other age groups as therapy needs to be intravenous and some antimicrobial classes are contraindicated, however delay in establishing appropriate therapy is even more dangerous.

The population structure of *Kpn*, similar to other opportunistic pathogens that thrive in the environment, is highly complex(5). Understanding long-term population trends of clinically relevant lineages at both a local and global scale, and which antimicrobial resistance plasmids and elements they encode, is potentially relevant to guiding treatment, infection prevention strategies and vaccine development. However, few longitudinal studies of *Kpn* population have been published so far, and the majority have come from collections established in high-income settings or South East Asia(6–10), whilst data from Africa and South America remain sparse. Low-income countries in sub-Saharan Africa are particularly underrepresented in 'global' collections of pathogen genome data, despite often carrying the highest burden of disease(2).

Whilst *Kpn* has a global distribution, there are clear geographical concentrations of dominant lineages and resistance genes, likely reflecting local antimicrobial pressure, patient populations, and a range of other determinants. Developing a more representative picture of the global epidemiology of *Kpn* is relevant to the development of vaccines targeting *Kpn*, to ensure that the lineages and associated LPS O-antigens and capsular polysaccharide antigens most prevalent in countries with the highest burden of disease are targeted. A better understanding of the genomic epidemiology of *Kpn* in low-income countries is also crucial to advocate for access to antimicrobials, in particular newer agents, which is often profoundly limited by cost, leaving isolates that can be readily treated in high- or middle-income countries *de facto* untreatable. Further, knowing at a facility level where problematic hotspots of transmission are within a hospital setting can help focus infection prevention and control (IPC) efforts.

To address this urgent need for a better understanding of the *Kpn* population in sSA, we performed whole-genome sequencing on all bloodstream isolates phenotypically identified as *Kpn* from Queen Elizabeth Central Hospital (QECH) in Blantyre, Malawi, which has undertaken sentinel surveillance of bacteraemia and meningitis in partnership with the Malawi-Liverpool Wellcome Programme (MLW) since 1998(3). Our analyses include all isolates, irrespective of resistance profile or patient demographic, from 1996 to 2020, with no change

in sampling strategy from 2000 - 2020. We report the shifts in resistance profile and population structure over time and highlight the relevance to consider trends over time and not just total numbers for considerations of new drug regiments or vaccine target selection.

Results

The QECH bloodstream isolate collection

QECH is the main referral hospital for the southern region of Malawi and the main hospital for Blantyre, the second largest city of Malawi (population 800,024 at the 2018 census). It has approximately 1,000 beds and receives around 10,000 adult medical and 30,000 paediatric medical admissions per year. A quality assured diagnostic microbiology service was initially implemented in partnership with MLW at QECH in 1998. Laboratory blood culture was initially manual, and an automated system was implemented in 2000 (BacT/ALERT, Biomerieux). The diagnostic microbiology service expanded to culture cerebrospinal fluid (CSF) from 2002. During the study period, all adult and paediatric medical patients with suspected sepsis or meningitis were eligible for blood and/or CSF culture at the discretion of the admitting physician, and where possible before the administration of antimicrobials. Bacterial isolates were identified by standard techniques detailed in the methods.

At the start of the period, first-line treatment against bloodstream infections was crystal-penicillin and chloramphenicol although gentamicin was used in neonates instead of chloramphenicol. This began to change with the introduction of the 3rd-generation cephalosporin ceftriaxone from 2004, which was rapidly adopted as a first line agent by the Department of [adult] Medicine and more slowly by the Department of Paediatrics, which continues to use penicillin and gentamicin in the management of neonatal sepsis, as per WHO guidelines(11).

Members of the order Enterobacterales including *Kpn* were identified by API biochemical index (Biomerieux, France). *Klebsiella pneumoniae* is known to comprise a species complex including *Klebsiella pneumoniae subsp. pneumoniae* as well as several closely related species like *K. variicola* and *K. quasipneumoniae* and their respective subspecies(12). These can only be identified to species complex level via API, and we therefore refer to all members of the *K. pneumoniae* species complex as *Kpn* and specify when referring to specific (sub)species (Fig 1A).

Dramatic increase in Kpn cases from 2016

The crude frequency of *Kpn* cases was high at the start of the period with 121 cases in 1998, although this overlapped with a year of active surveillance of febrile adult patients(13) and we thus focused longitudinal analyses from 2000 onwards. The number of cases from 2000 - 2015 was stable with a mean of 44 cases per year (Fig 1A). This increased to a mean of 187 cases per year from 2016-2021 with a peak of 279 in 2019.

Neonatal infection with *Kpn* accounted for 39.3% of all *Kpn* infections, even though only 11.2% of all blood cultures were taken from this patient group. The overall rate of *Kpn* isolation was 24.9 per 1000 blood cultures for neonates, 4.3 per 1000 blood cultures for other patients (rate ratio of neonates vs non-neonates 5.11 [95% CI 4.66 - 5.61]). The increase in *Kpn* cases post 2016 was driven by an increase in neonatal infections (S1A-1G Fig); there were peaks in the *Kpn* blood culture isolation rate in neonates in 2008, 2014 and 2017 which suggest outbreaks or ongoing transmission on the neonatal unit (Fig 1B), and our genomic analyses further

support hospital transmission given the close genetic relationships of the respective isolates. Though the blood culture positivity rate was increased for most age groups in the period from 2016-2021, for neonates this increase was particularly marked (33.5 per 1000 blood cultures in the period 2016-2021 compared to 15.4 per 1000 blood cultures in the period 1998-2015; S1A-1D Fig). As well as the peak in neonatal cases, there was also a much smaller peak in the adult age group, which shifted from the 30-35 year age group in the period from 1998-2015 to the 35-40 year age group in the period from 2016-2021 (S1E-1H Fig).

Increase in *Kpn* numbers driven by temporally defined expansions of sequence types (STs)

We retrieved 1313 isolates not previously sequenced from the QECH *Kpn* collection, 1302 of which could be recovered for growth and yielded enough DNA for sequencing; and 1156 of these passed all quality control steps for further analysis (details see methods; S2 Fig). In addition, we included 217 samples from previously published studies of this collection(14–16), resulting in 1373 sequenced genomes analysed in this study (Fig 1C).

The majority (1274/1373, 92.8%) belong to *Klebsiella pneumoniae* subsp. *pneumoniae* with only very small proportions representing other (sub)species, including *K. quasipnemoniae* subsp. *quasipneumoniae* (10/1373, 0.7%), *K. quasipneumoniae* subsp. *simillipneumoniae* (46/1373, 3.4%) and *K. variicola* subsp. *variicola* (43/1373, 3.1%). Based on phenotypic microbiology data (Fig 1A and 1B) we observed a drastic increase in isolates from neonatal infection in 2016/2017 which is reflected in our sequenced collection.

Assessing diversity of STs over time highlights that the increase in *Kpn* infections represents repeated increase and subsequent decrease in numbers of different multilocus sequence types (Fig 2A). Whilst the earlier years show a diverse composition of STs with small numbers of each (n < 15 total), the increase in total *Kpn* numbers is mainly driven by one or two major STs, with these peaks spanning approximately 1.5 to 2 years each. We furthermore observe that after the increase of isolates of the major STs, each of them reverts to their original very low numbers (Fig 2A).

A previous study(15) highlighted an outbreak of ST340 across the neonatal unit which is reflected here, where it comprises 31 out of 105 isolates in 2013 & 2014 (29.5%; Fig 2A). We then see an increase of the numbers of ST14 (19 / 44 in 2015; 43.2%) which is present at a lower proportion before and after this period; followed by an increase in ST25 (13 / 87 in 2016; 14.9%) which again is present at lower proportions before and after. The year 2017 then saw a steep increase in *Kpn* infections overall; mainly driven by an expansion of ST39 which comprises more than half the samples at the time (100/188 in 2017; 53.2%). This clone however again retreats back to low numbers, with a few lineages dominating, in particular ST1552 (35 / 208 in 2018; 16.8%) and ST35 (40 / 208 in 2018; 19.2%). The year 2019 then sees another increase in ST14 (61 / 236 in 2019; 25.8%) as well as ST15 (30 / 236 in 2019; 12.7%), which is a known high-risk clone predominantly found in South East Asian collections where it has been linked to an early spread of carbapenem resistances via bla_{NDM} genes(17)(18).

Isolation of *Kpn* from neonatal and paediatric wards account for observed changes, whilst adult infections remain stable over time

The patterns over time differ significantly between wards (Fig 2B), suggesting the increase in *Kpn* infections could represent a pattern of repeated ward-specific outbreaks, rather than a generalised increase across all wards (Fig 2B). This pattern of ward-focused increases in isolate numbers was repeatedly observed on the Chatinkha neonatal unit and the surgical PICU/HDU, where most of the changes in numbers of isolates occur. An indication of local outbreaks is further supported by analysing the major STs, which increase at defined times in specific wards (Fig 2B and 2C; S3 Fig). To investigate the impact of closely related isolates (<5 core SNP distance) that are likely within-hospital transmissions on the number of infections overall, we analysed a core genome tree for each of the three species and identified closest-neighbour SNP distances in a root-to-tip approach(19) (Fig 3). This highlighted that most highly clonal lineages (defined as <5 SNPs) are derived from *K. pneumoniae* subsp. *pneumoniae* (Fig 3A) representing 402 of 1165 isolates as a conservative estimate based on whole-species analyses, with 43% of the *K. quasipneumoniae* subsp. *similipneumoniae* (20 of the 46 isolates in 6 lineages, Fig 3B) and 25% of the *K. variicola* isolates (10 of the 40 isolates of this species, Fig 3C); and no <5 SNP related isolates of *K. quasipneumoniae* subsp. *quasipneumoniae*.

Despite removing the single-isolate STs for this analysis, our dataset still contains a high number of *K. pneumoniae* subsp. *pneumoniae* STs, which results in a highly diverse dataset. As this diversity has potential to impact on tree resolution depending on reference used(20), we investigated three key STs in more detail. We focused on ST35, ST39, and the closely related ST14 and 15 together, whose pattern of accumulation in a ward and year strongly indicates multiple hospital-associated transmissions (S3 Fig), in more detail by mapping against strain-specific reference genomes (Fig3A). ST39 and ST35 are two key STs when considering the per-year breakdown of isolates per ward for the neonatal unit (Chatinka; S3 Fig); whilst ST14 and ST15 were identified in high numbers in the PICU/HDU in the paediatric surgical ward (S3 Fig). Here, we can observe that 77% of the isolates (51/55 ST35, 110/136 ST39, 182/254 ST14-15) have their nearest neighbour within the threshold of 5 SNPs; representing 100% (17/17) of ST35 and 88.9% (80/90) of ST39 from Chatinka and 88% (57/65) ST14 and 15 from the PICU/HDU (Fig3A). This high relatedness of isolates strongly supports within ward transmission and the argument that *Kpn* is a healthcare associated cause of neonatal sepsis. As patients are regularly transferred from Chatinkha to the PICU/HDU if surgical procedure is required, these wards are furthermore already linked by a clear transmission path via patients, and several potential between-ward transmission events are indicated in particular in ST35 and ST14/15 (Fig 3D, F).

Over the period from 2015-2018 there was an increased focus on performing blood cultures to ensure every neonate on Chatinkha nursery that met the clinical criteria for suspected neonatal sepsis was included, which may explain some of the large increase in *Kpn* cases in 2017 on the nursery. Establishing optimal IPC has been an ongoing challenge(21); specifically, handwashing has been difficult due to intermittent water supply and a lack of access to soap or hand sanitiser and there are limited numbers of cots, so neonates have often shared them. In response to the large increase in *Kpn* cases there was renewed focus on IPC; hand sanitizer and soap procurement was prioritized, deep cleans of the ward environment were performed regularly, there was increased cleaning of ward surfaces, ward traffic was limited, wooden cots are being replaced with easier to clean plastic ones, the number of neonates sharing a cot was reduced and oxygen delivery devices are being replaced with single use items.

Antimicrobial resistance and implications for treatment

Phenotypic resistance profiles of the strains, where available, mirrors the very concerning trends reported previously for *Kpn* in QECH(3) (Fig S4, Table S4). Resistance against the frontline treatments is widespread; a particular concern is a high proportion of isolates with ceftriaxone, gentamicin and cotrimoxazole resistance from 2012 onwards with very little change over time (S4A-C Fig; >75% of tested isolates were resistant to these agents each year 2012 – 2020). We notice some variation in co-amoxiclav (augmentin: amoxicillin + clavulanic acid) susceptibility profiles but from 2019 again show >75% of tested isolates as resistant (S4A Fig). Importantly, we notice that a trend of increased numbers of chloramphenicol susceptible isolates (Fig 4B, S4G Fig), which has previously reported in Malawi(3), is also explicit in our *Kpn* collection in particular in recent isolates (38.7%, 51.5%, 72.2% isolates sensitive for 2018-2020). Presence of *catA1* or *catA2* agree well with phenotypic resistance whereas presence of *catB4* did not as shown in Fig 4B; where we see a large number of isolates with sensitive phenotype, no *catA1/2* but *catB4*, whereas only a small fraction of isolates has resistant phenotype, no *catA1/2* but *catB4* where we hypothesize a different mechanism is likely driving the resistant phenotype. For isolates with resistant phenotype and encoding both *catA1/2* and *catB4* the resistance is presumed to be driven by *catA1/2*. This agrees with *catB4* representing a non-functional truncated variant of *catB3* as was recently reported (Graf et al. 2023). We furthermore note a sensitivity of 50% or higher in the majority of years measured in ciprofloxacin (6/9 years from 2012 to 2020, including 2017, 2018 and 2020; S4E Fig, Table S4), which together with chloramphenicol thus represent two antimicrobials that might thus be valid treatment options.

Resistance to 3GC was widespread (83% resistance to cefpodoxime/ceftriaxone by 2020; S4A Fig) and almost exclusively driven by acquisition of the $bla_{CTX\text{-}M\text{-}15}$ gene, which continuously outcompetes other ESBL-encoding genes that are occasionally present in the *Kpn* population, but which are only sporadically detected (Fig 4A, Table S2). The number of plasmid replicons per isolate stabilised after the first increase in numbers around 2015, highlighting that a few large AMR plasmids are likely responsible for the majority of AMR-isolates (S5C, S6A Fig, Table S3). Strains with a moderate number of plasmid replicons dominate the population (0-5) and this remains stable over time, with only a small number of isolates with more than five plasmid replicons associated that remains at similar numbers over time (S6A Fig). This indicates that the change to a predominantly $ESBL^+$ population was the key event, with the population then remaining stable over time with respect to plasmid diversity and numbers of resistance genes (S6B Fig).

Carbapenem resistance genes are still only rarely observed; we identified two samples encoding for $bla_{NDM\text{-}1}$; one in a ST39 in 2016 and one in a ST45 in 2019, perhaps unsurprisingly as carbapenem usage is still comparatively unusual. We also investigated resistances against other potential alternative treatment options colistin, fosfomycin and tigecycline. Resistance to these agents are all at low levels, with only one acquired fosfomycin resistance gene detected (*fosA*) and none against colistin or tigecycline predicted in our dataset, and one isolate with predicted mutations that can confer resistance to colisin (*pmrB*), indicating that these might be viable alternatives for $ESBL^+$ infections at QECH, although tigecycline is contraindicated in children under 8y (Table S2) and neither agent may not be ideal treatments for bacteraemia(22,23). Tigecycline resistance is often conferred by upregulation of transport proteins and thus challenging to predict and has also been observed during within-patient

evolution(24–26); we could however not detect the described plasmid-derived antimicrobial resistance genes *tetX* in our collection(27).

Based on AMR phenotype identified by disc diffusion, amikacin remains a viable alternative to gentamicin, although there are concerns over blood brain barrier penetration in cases of meningitis and renal toxicity in settings where therapeutic drug monitoring is not possible. Phenotypic data are corroborated by a more detailed analysis of the aminoglycoside AMR genes (Fig 4C and 4D). In total 90.5% of all isolates (1243/1373) are predicted to encode one or more aminoglycoside resistance genes, however, the majority are predicted to confer resistance to streptomycin (ANT3) and gentamicin (AAC(3)-II; ANT(2")-I; Fig 4D). Only a fraction of these strains (43.2%, 593/1373) encoded one of the genes predicted to confer resistance against amikacin (AAC(6')-I; Fig 4C). Aminoglycoside resistance is known to be challenging to predict(28); whilst phenotypic gentamicin resistance testing matches very well with the predicted profile based on acquired genes, the presence of AAC(6')-I does not by itself confer resistance in a large number of isolates, with only 12 isolates encoding for either of these, representing 2%, actually showed a resistant phenotype (Fig 4D; Table S4).

Potential vaccine target stability

The *Kpn* polysaccharide antigens are of high interest as vaccine targets; in particular the large capsular antigen. As expected, we observed a very high number of different K-loci (90 predicted), as well as a diverse set of O-loci (15 predicted), the latter are known to be far less diverse in *Kpn*(29). Recent findings furthermore showed that *wbbZ,* encoded on the chromosome outside the O-locus, contributes to the diversity of the O-Ag structure displayed by *Kpn* and thus two isolates with the same O-locus can still result in different O-types based on the presence or absence of this gene. Immunologically relevant differences in O-Ag thus requires the prediction of the O-Ag locus and *wbbZ*(30), so we added this to our O-Ag type predictions (Fig S7A-D) where we thus distinguish between O1a and O1b as described by Kelly *et al*(30). Reflecting the trend observed with the ST types, we observe temporal variation in prevalence of O-types and K-types (Fig 5A and 5D, S7A Fig). The non-ESBL population again displays a much greater diversity for both types of antigen (Fig 5B and 5E, S7B Fig), whilst the ESBL population is more biased towards dominant O- and K-types (Fig 5C and 5F, S7C Fig).

Beyond the dominant O1b (821/1373; 59.8%) and O2afg (184/1373; 13.4%) predicted O-antigen types, the infrequently observed type O4 comprises a substantial fraction of isolates (90/1373; 6.6%); and 35 isolates have a prediction accuracy as "Low" or "None" which can indicate either low assembly quality or novel O-Ag loci and types. In addition, we note two isolates with unknown O-Ag locus encoding for additional *wbbY* and *wbbZ* genes, indicating that more variation might be within possibly novel O-Ag loci depending on additional enzymes encoded in the chromosomal backbone (Table S5). We furthermore note 8 isolates with O1/O2 loci encoding for only *wbbZ* (not *wbbY*); whilst the recent work showed no structural differences in the O-Ag if *wbbZ* alone was added, we cannot exclude other carbohydrate enzymes on the chromosomes of our isolates that might lead to additional different structures if combined with *wbbZ* (Table S5). Of relevance, we also observe variation of O- and K-type combinations within the same sequence type (Fig 5G and H, S7D Fig) and with each other (Fig 5I, S7E Fig).

Considering hypothetical vaccine strategies and constructs, an efficacious maternally administered quadrivalent O-antigen based vaccine (O1a, O1b, O2afg, O4; including O1a as O1b-enconding isolates can switch between O1a and O1b(30)) would be expected to protect against up to 89% of cases of neonatal sepsis in Blantyre (345/388), whereas a K-antigen based vaccine would require 22 K-antigens to be represented to protect against an equal number of cases of neonatal sepsis when taking the entire timespan of our study into account (KL10, KL102, KL103, KL104, KL106, KL108, KL117, KL142, KL15, KL16, KL17, KL19, KL2, KL23, KL25, KL30, KL43, KL48, KL51, KL52, KL62, KL9); a quadrivalent K-antigen based vaccine (KL2, KL25, KL30, KL102) would cover 58% (198/388) of cases. We furthermore observe no regular cycling between a specific set of K-types, but rather that different K-types regularly emerge associated with major clones (Fig 5). Given current knowledge of >180 predicted K-types(31) and >70 determined structures so far(32–34) (likely an underestimation), and the propensity of the K-locus to recombine, it will be very challenging to predict how long a K-type vaccine would remain effective.

Trends of resistance vs virulence determinants

Whilst most *Kpn* isolates are opportunistic pathogens with poorly understood virulence profiles, the hypermucoid phenotype is a known virulence determinant arising from hyperproduction of capsule and causing a different clinical infection type, i.e. liver abscess and meningitis in otherwise healthy individuals, often progressing very rapidly and with poor prognosis. The capsule hyperproduction is usually driven by acquisition of a virulence plasmid encoding for the capsule polysaccharide upregulation factor *rmpA2* as well as several siderophore systems (aerobactin, salmochelin, colistin) that enable iron scavenging. This phenotype is of particular concern given the recent convergence of hypervirulent with multidrug resistant isolates. Whilst we note hypervirulent isolates present (defined here as kleborate virulence score >1, i.e. encoding virulence factors in addition to the often chromosomal yersiniabactin) from the very early sampling times on in our collection (S8A,B Fig, S5B Fig), these remain sporadic; in contrast to multidrug resistant isolates (S8C,D Fig) which are absent at these early sampling times but rapidly become the majority of isolates from 2007/8 onwards (Fig S8D). Whilst we note a total number of >10 isolates for both ST86 and ST268, two STs carrying the *rmpA2* gene and widely observed as hypervirulent phenotypes, these occur in different years and wards, highlighting that hospital-derived transmission is highly unlikely (Figure S8B). Whilst it is of concern that we observe some of these potentially hypervirulent isolates (encoding for *rmpA2*) also encoding an ESBL gene (22 total; 9 of these ST86, no ST268) there seems to be no selective advantage; the first isolate carrying this combination is observed in 2010 but a maximum of 4 isolates per year are observed until 2020. This small number of putative hypervirulent isolates is in accordance not only with bloodstream isolates in general but also with a study comparing carriage with infectious isolates from adults in Blantyre, where only a very small number of isolates encoding virulence determinants were identified, all of them in infectious isolates(35,36).

Discussion

Neonatal sepsis is responsible for a high proportion of deaths in children under 5 in sSA, with *Klebsiella* by now being both the most prevalent gram-negative pathogen and increasingly hard to treat. We have analysed a longitudinal collection of routine diagnostics invasive (bloodstream and CSF) isolates of *Kpn* spanning from 1998 – 2020, during which time AMR increased drastically. We determined longitudinal patterns of patient demographics, *Kpn* population trends, vaccine targets, resistance elements and remaining treatment options. This

exceptional resource, not only for sSA but also world-wide, enables us to understand the dynamics of *Kpn* over an exceptionally long time scale where the sampling was not biased towards resistance types. In particular, it allows us to see the dynamics before, during, and after the significant expansion of ESBL producing *Kpn* at maximum resolution; and highlights the need to consider not only major types based on total numbers but the temporal dynamics of vaccine targets such as cell surface polysaccharides.

Kpn initially showed a bimodal age distribution in patients at QECH(3) and there were concerns that increasing life expectancy would lead to a higher number of adult infections; however, this trend has yet to be seen, and adult bloodstream infections were significantly reduced overall following the widespread distribution of antiretroviral therapy(37). The stark increase in infections from 2016 onwards is almost exclusively seen in neonates and children where ward-restricted clonal expansions that are subsequently replaced with other successful clones. This suggests that context appropriate, effective, scalable infection prevention and control (IPC) targeted towards the most vulnerable patient groups in the hospital and a better understanding of *Kpn* transmission routes and reservoirs would prevent a large proportion of the cases caused by *Klebsiella* bloodstream infection. Such approaches have the potential to protect vulnerable babies against all healthcare associated infections rather than just one pathogen and are available now, unlike a vaccine strategy that will take years to develop and implement without guarantee for success.

The main sequence types in our study were ST39, ST14 and ST15. While ST39 and ST14 are not necessarily the most prevalent ones observed globally, ST15 is widely reported in SouthEast Asia causing clonal outbreaks whereas it showed a different pattern in our setting driving smaller expansions but also as a lot of sporadic isolates occurring over time. This highlights the need for more data from sSA to understand the complex diversity of this organism in high-risk areas, as not only the ST composition, but also their patterns, differ between settings. The two key insights we gain from the population pattern is that the lineages continue to replace each other as the dominant clone over longer time periods, and that in our setting, this expansion is mainly driven by an increase in cases in neonatal patients. Most of the major sequence types are present at a steady, much lower proportion, before and after their respective clonal expansions, emphasizing the relevance of understanding what leads to the wave-like pattern of clonal expansions in *Kpn* populations.

This temporal variability also favours prioritisation of IPC strategies to prevent the introduction of new lineages. It may be that increased IPC efforts in response to ward outbreaks contribute to or are the single cause for the sudden stops of clonal expansions. We have little understanding of the ability of *Kpn* to survive on abiotic surfaces, which represents one of the key differences between HIC and LMIC settings, as the latter often use wooden furniture and bristled plastic tubing due to limited resources, which might lead to different *Kpn* lineages being successful. A similar question remains open regarding agents used for disinfection, as resource-limited settings often rely on basic soap agents and use unsafe water, whilst HIC routinely use industrially produced, quality controlled disinfectants.

Our data allow insight into AMR patterns in our setting. Ceftriaxone resistance, almost exclusively due to the ESBL-encoding $bla_{\text{CTX-M-15}}$, is widespread. Whilst carbapenems are rarely available and resistance is already emerging (38,39), several treatment options are still viable, such as amikacin. Our analysis also emphasizes again that aminoglycoside resistance

is notoriously hard to predict and might be impacted by the genetic background (for example efflux pumps) and not only acquired resistance genes, and should therefore at this stage not be ruled out based on genomic data but confirmed via phenotyping. Most isolates were phenotypically resistant to gentamicin; whether this is consequent on the predicted resistance genes, or whether other mechanisms drive this resistance is important to determine to improve the value of genotypic predictions for aminoglycosides. It will remain to be further investigated whether this is due to other interfering genes or low expression of the resistance gene and might be an indication of heteroresistance in the population which needs to be carefully considered(40). We furthermore observe a trend in re-emergence of chloramphenicol sensitivity, putatively in part driven by the spread of a non-functional CAT copy (Graf et al. 2023). Continuing to monitor re-emergence of susceptibility to agents no longer in use is of high relevance to allow timely re-introduction of drugs and can give us insights into the potential of rotating antimicrobial regiments over longer time spans to allow sensitivity to re-emerge.

There was little resistance predicted and observed to colistin, a last-line treatment option, in QECH *Kpn* isolates. Especially the lack of mobile resistance elements is encouraging, but also stresses the urgent need for better management of animal health as colistin is widely used, knowingly and as part of food products farmers are not necessarily aware of, in the increasing mid-size farming industry(41). Another exciting new development are initial results of combining fosfomycin and flomoxef(42); we observe only sporadic fosfomycin resistance mechanisms and very little spread of *ampC* genes, the only non-carbapenemases that would deactivate flomoxef, although we note that there are also counterindications for fosfomycin use in invasive infections.

As extensively resistant bacterial pathogens emerge and spread, there is an increasing urgency to develop vaccines to break the arms race between new antimicrobials and new resistance mechanisms. In this case there is interest in developing maternally administered *Klebsiella* vaccines to prevent neonatal sepsis. A target of high interest are the extracellular polysaccharides that can be typed based on their respective operons, which are assigned different K-types (capsular antigen) and O-types (the lipopolysaccharide O-antigen chain). The K-antigen shows a very high diversity; and whilst we can identify several dominant types in our study, the long timeframe of our study highlights the importance of considering the dynamic nature of K-antigens when considering vaccine design. Here, temporal fluctuations detected just in Blantyre strongly suggest that a K-type vaccine would become ineffective very rapidly given the frequent changes in dominant K-type, which might be driven by phage pressure in the environment and thus highly unpredictable. It is also unclear if successful vaccination against the major polysaccharide types would lead to a degree of cross protection against non-targeted types or whether lack of such cross protection would facilitate rapid emergence of infections caused by vaccine escape lineages.

In contrast, there are far fewer O-antigen types, making this a more attractive target. Recent efforts have compared several large datasets with O- and K-typing either determined by PCR or WGS(43); and identified that a conjugate vaccine targeting O1v1/2 (~serotype O1)(31), O2v2 (~serotype O2afg)(31) and O3b (~serotype O3b)(31) has the potential to prevent the majority (>70%) of cases of *Kpn*-BSI. For the most at-risk population in our setting, the neonates, this would cover 87% of cases (338/388) assuming O1a and O1b were included into O1 and no other subtypes were discovered(30); importantly it would however not have

prevented the ST340 outbreak in the neonatal unit previously described(15) (O4). It thus seems of relevance to consider including the currently low-abundance O-antigens as well, in particular given the interchangeability for O-types within the same STs. Reports from other highly understudied settings also report other main O-types in hospital isolates, e.g. O5 in a study in Pakistan(18,44) or a diversity of O-types in the Caribbean where this O-Ag vaccine would only cover 62% (162/260) (45). Selection pressure might not directly drive evolution given that *Kpn* is not human restricted but could likely select for strains with currently less abundant O-types to increase in prevalence. Further, there remains an open question of how accessible the O-antigens are to antibodies, which a recent study calls into question(46); we currently have very limited understanding what drives the extent of capsule expression per cell in non-hypermucoid isolates and how different stages of the infection impact this. It is known from *in vitro* studies that capsule expression even in the same experimental setting can differ widely within clonal lineages, which also impacts the bacterial ability to escape the complement system(47).

Our data, and the recent identification of a novel form of the long-studied O1 antigen type(30,48–50), highlights that we have very little understanding of the relevance of the O- and K-types on the success of *Kpn* lineages, whether any additional genes on the chromosomal backbone can further modify the structures and lead to different serotypes with the same O-locus, as well as whether any O-/K-combinations are more or less favourable. Closing such knowledge gaps will be central to design of vaccines as we also have little understanding for the propensity of sequence type combinations with specific vaccine target serotypes such as capsule or O-antigen. A more solid understanding of O-/K-type variations, both regarding their biochemistry and whether they contribute to epidemiological success, is crucial given the high interest to use these as targets for maternally administered vaccines against neonatal sepsis.

This highlights that we have very little understanding of the relevance of the O- and K-types on the success of Kpn lineages, whether any additional genes on the chromosomal backbone can further modify the structures and lead to different serotypes with the same o-locus, as well as whether any O-/K-combinations are more or less favourable. A more solid understanding of O-/K-type variations, both regarding their biochemistry and whether they contribute to epidemiological success, is crucial given the high interest to use these as targets for maternally administered vaccines against neonatal sepsis.

In summary, our study provides a unique insight into long-term patterns of *Kpn* hospital isolates from a low-income setting, spanning across a time period where the number of MDR and ESBL producing infections increased rapidly across the globe. We show that whilst STs known as high-risk in other settings are common, these do not always show similar success in different settings; and we observe a significant numbers of isolates from otherwise rarely observed STs, causing hospital infections. This again emphasizes the importance of including the relevant settings into studies aiming to identify treatment or vaccine targets as we currently have poor understanding of what makes an ST a high-risk lineage in each setting at a given time. Our data also emphasizes the urgent need to consider temporal variation when considering treatment or vaccine targets, as some of the main vaccine targets show strong temporal fluctuations; the most prevalent type now circulating is unlikely to cause the next outbreak. The close relatedness of isolates from the same ward in the same time frame strongly indicate that the majority of infections are hospital-acquired, so whilst vaccine

development and alternative treatment options are highly relevant; at this stage, focused efforts to improve IPC are key to reducing *Kpn* disease burden now(21).

Materials and Methods

Isolate collection

QECH is a government run tertiary referral centre for the southern region of Malawi, which provides free care at the point of delivery. It serves a population of 800,000 for the area of urban Blantyre (Malawi census 2018). There are adult and paediatric A&Es, medical and surgical wards. There is a tertiary neonatal unit on site and a separate paediatric surgical hospital which opened in 2017 (Figure 2C; PICU, Paediatric HDU, Paediatric surgical ward).

Microbiological sampling and processing

Routine, (ISO15189 accredited since 2019) diagnostic blood culture services have been provided to the medical and paediatric wards by the Malawi-Liverpool-Wellcome Programme since 2000. For adults 7-10mL of blood were collected from all patients admitted to the hospital with fever (axillary temperature >37.5°C) or clinical suspicion of sepsis, severe sepsis, or septic shock. Sepsis, severe sepsis, or septic shock were suspected in patients with tachycardia (≥90 beats per minute), hypotension (systolic blood pressure <90 mm Hg), tachypnoea (respiratory rate >20 per minute), or delirium. 3–10 mL of blood was taken from children with non-focal febrile illness who tested negative for malaria, who were severely ill with suspected sepsis, or who failed initial malaria treatment and remained febrile. For adults and children with clinical suspicion of meningitis (temperature >37.5°C with seizures, headache, abnormal behaviour or meningism) a lumbar puncture was also performed and the sample sent for CSF analysis. For neonates with clinical signs of sepsis or meningitis (temperature >37.5°C, or other signs of clinical deterioration such as reduced activity, increasing work of breathing or respiratory rate >60, heart rate >180 or <100, and unexplained seizures) a blood culture and CSF sample was collected. Afebrile patients were unlikely to have had blood or CSF cultures unless the clinical suspicion for sepsis or meningitis was high.

For blood cultures, samples were collected using aseptic methods and inoculated into a single aerobic bottle (BacT/Alert, bioMérieux, Marcy-L'Etoile, France). These were incubated using the automated BacT/Alert system (bioMérieux, France) since 2000. Before which, they were cultured manually. Samples that flagged positive were Gram stained and Gram-negative bacilli are identified by Analytical Profile Index (Biomérieux). CSF samples were processed for cell count and biochemistry before manual culture. Antimicrobial susceptibility testing was determined by the disc diffusion method (Oxoid, United Kingdom) using AST breakpoints that were current at the time, initially following the relevant version of the British Society of Antimicrobial Chemotherapy's guidelines until 2018, after which time EUCAST methods and breakpoints were introduced. Organisms which showed reduced inhibition when exposed to Ceftriaxone 30mg discs, Ceftriaxone 5mg discs, or Cefpodoxime 10mg discs were classified as resistant to Ceftriaxone. Intermediate isolates were classified as resistant to ceftriaxone. Details for the identification of other organisms are described elsewhere(3). *Klebsiella pneumoniae* isolates were stored at -80°C on microbank beads.

Whole genome sequencing and QC

All *Klebsiella pneumoniae* isolates that could be recovered from the start of the MLW archive to April 2020 were thawed and incubated on MaConkey's media for 18-24 hours at 37°C. Plates with pure growth then had a single colony pick taken and inoculated into 15 ml of

buffered peptone water for 18-24 hours at 37°C. These samples were then centrifuged and the supernatant was discarded. The pellet was then resuspended in buffer and the DNA was extracted using the QIAsymphony machine and QIAsymphony DSP kit with onboard lysis, according to the manufacturer's instructions. Quality control was done using Qubit and samples with a DNA volume of less than 200ng were repeated. DNA was sequenced at the Wellcome Sanger Institute on the Illumina HiSeq X10 instrument (Illumina Inc., United States) to produce 150 bp paired end reads.

Eleven out of 1313 samples failed the Sanger-internal sequencing threshold for minimum coverage for assemblies and were thus removed from further analyses. Quality control on all remaining isolates was performed using species confirmation via Kraken v0.10.6(51) to identify contaminants (incorrect species). Any species assignment other than *Klebsiella pneumoniae*, *Klebsiella variicola*, unclassified and synthetic construct were considered contaminants; the latter two categories frequently derive from plasmid sequences in our experience. All isolates with higher than 10% of other species were removed as contaminated (S2A Fig). Similarly, isolates with a ratio of heterozygous SNPs over 5% were removed as likely contaminated by a related strain (S2B Fig). Further quality control used the assemblies, where all isolates resulting in more than 1000 contigs were removed as likely contaminated by other species or strains, or of very low quality (S2C Fig). *De novo* assembly of the remaining 1156 genome sequences was performed using SPAdes v3.14.0(52) incorporated into the bacterial genome assembly pipeline described previously(53). Annotation of all assemblies was performed using prokka(54) v1.14.5.

Genome analyses and phylogenetics

For comparative analyses, isolates from recent studies at QECH were included, as they are part of the bloodstream collection(14–16). Details of all sequences used in the analyses are given in supplementary table 1. Resistance gene prediction and plasmid replicon types were performed using ariba(55) (v2.14.6) and the srst2-argannot(56,57), CARD(58) and plasmidfinder(59) databases (download date 14.02.2022). To assess aminoglycoside resistance; ANT3 is by ariba as AacAad_AGly, AadA4_5_AGly, AadA_AGly, AadA_AGly_2, AphA6_AGly; AAC(3)-II is grouped by ariba as Aac3_IIa_AGly; ANT(2")-I is grouped by ariba as AadB_AGly and AAC(6')-I is grouped by ariba as AacAad_AGly_3.

Prediction of sequence types, (sub)species and *Klebsiella*-specific virulence factors were performed using kleborate v2.1.0(60) and the –r setting; the results are given for all isolates in Table S1. O- and K-types were predicted with kaptive(31) (v2.0.6) which incorporates O-loci as well as O-serotype prediction(29,31,61); and we added an additional option of *wbbZ* as extra gene to distinguish between O1a (O1/O2 + *wbbY*, no *wbbZ*) and O1b (O1/O2 + *wbbY* + *wbbZ*) as described recently(30), adding *wbbZ* as implemented in kaptive versions before v2(61) and expanding the .logic file to distinguish O1a/b.

Core genome SNPs for *K. pneumoniae* subsp. *pneumoniae* isolates were determined using snippy(62) (v4.6.0; https://github.com/tseemann/snippy) with the *Klebsiella pneumoniae* subsp. *pneumoniae* NTUH-K2044 genome (AP006725.1)(63) as reference. Recombinant regions were removed using gubbins(64) (v3.2.1) and the FastTree(65) option ('-tree-builder fasttree'), given the very high number of samples, as implemented in gubbins; and only isolates from sequence types with more than one isolate in our dataset were included in the analysis to further reduce compute time. The same strategy was used for *K. quasipneumoniae*

and *K. variicola* using strains ATCC 700603 (CP029597.1)(66) and At-22 (CP001891.1)(67) as reference, respectively, where all isolates were included and default options for gubbins were used (RaxML for tree generation, given the much smaller sample numbers). A snp-only alignment was generated from the recombination-free gubbins output using snp-sites(68) v2.5.1 with the -c option (selecting for ACGT-only) for *K. quasipneumoniae* and *K. variicola*; for *K. pneumoniae* subsp. *pneumoniae*, given the high variation of isolates, the -c option was too stringent (resulting in a <200 bp alignment) and we thus used snp-sites to isolate all snps but without selecting for ACGT-only sites.

The snp-only alignments were used as input for iqtree(69) (v1.6.12), the number of constant sites was determined with snp-sites -C function from the original full alignment and provided to the '-fconst x,x,x,x' iqtree function. The model was determined with the integrated ModelFinder(70) function '-m TEST' resulting in GTR+F+I+G4, GTR+F+I+G4, TVM+F+I+G4, for the alignments from *K. pneumoniae*, *K. quasipneumoniae* and *K. variicola*, resp. We set '-keep-ident' to keep identical data and '-bb 1000' for 1000 ultrafast bootstrap replicates. The resulting trees were then adjusted to SNP-length branch lengths using the python implementation of pyjar(71) (https://github.com/simonrharris/pyjar) as input for rPinecone(19). Closely related clusters taking their phylogenetic distance into account were determined with 5 SNPs as cut-off for minor and 10 as cut-off for major clusters. These were visualised with iTol(72) (v6.8) with visualisation files generated as implemented in rPinecone(19). For ST35, ST39 and ST14/15 (combining ST14 and ST15 isolates in one analysis) we performed ST-specific mappings of all respective isolates against references from these sequence types to gain higher resolution (CP041353.2(73), LR991401.1(74) and CP008929.1(75), resp.).

Pan-genome reconstruction, generation of a core gene alignment, and identification of neighbouring gene regions were derived via panaroo(76) (v1.3.3) using the '–clean-mode strict' setting. The phylogenetic tree based on the core gene alignment, with informative sites extracted with snp-sites as above, was calculated using iqtree(69) (v1.6.12) and the internal ModelFinder(70) function (GTR+F+ASC+G4) as above. All other metadata summary analyses and visualisation were generated using the ggplot(77) and ggtree(78) packages in R(79) unless indicated otherwise; some colour schemes are derived from or inspired by the MetBrewer (https://github.com/BlakeRMills/MetBrewer) and NationalParks (https://github.com/katiejolly/nationalparkcolors) R libraries. Composite figures were generated using the patchwork package (https://github.com/thomasp85/patchwork). An RMarkdown file and all underlying tree and alignment files are freely available online (zenodo doi).

Data availability

The RPackage contains all input scripts and files used to create the figures. All sequence data is available on SRA/ENA under BioProject PRJEB42462; specific accessions and additional data is provided in supporting table S1 which can also be found in the RPackage.

Acknowledgments

We would like to thank the Sanger Institute Pathogen Genomics group for expert computational support. We acknowledge and thank all the clinical staff and patients at QECH over the years for their contribution; and the MLW core microbiology team for expert support in storage and handling of the isolates.

Funding
We acknowledge funding from Wellcome (EH; grant 217303/Z/19/Z) and the BMGF (NAF and EH; grant INV-005180) as well as Wellcome funding providing core support for the Wellcome Sanger Institute (206194) and MLW (206454). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Ethics statement
Ethical approval for this study was granted by the University of Malawi College of Medicine Research Ethics Committee (COMREC) (P.11/18/2541) and isolates were shipped under a Nagoya Protocol compliant Access and Benefit Sharing agreement between the Government of Malawi and the Wellcome Sanger Institute.

Competing interests
The authors declare no competing interests.

Author contributions
Conceptualization: *EH, NAF;* Methodology: *EH, NAF;* Software: *EH, JML;* Validation: *EH, OP, AZ;* Formal analysis: *EH, OP;* Investigation: *EH, OP, AZ, PM, PS, ET, FG, RL, SL, JC, JML;* Resources: *SB, PS, ET, KK, NRT, NAF;* Data curation: *EH, OP, PM, RL, JC, JML, AZ;* Writing—original draft: *EH, OP, NAF;* Writing—review & editing: *all authors;* Visualization: *EH, OP, JML, NAF;* Supervision: *EH, CM, SL, KK, NAF;* Project administration: *EH, NAF;* Funding acquisition: *EH, NRT, NAF*

Figure legends:
Fig 1. ***Kpn* cases at QECH.** (**A**) Numbers of *Kpn* cases per year at QECH. Bars represent the crude frequency of *Kpn* infection for each year from 1998 - 2021, with the different colours representing the different age groups of the patients. (**B**) Blood culture positivity rate (per 1,000 blood cultures) of *Kpn* in neonates and the entire patient population.
Fig 2. **Uneven distribution of STs and cases at QECH wards indicate hospital-acquired infections.** (**A**) Main STs and (**B**) wards the samples were derived from over time. (**C**) A schematic map of QECH with the wards samples of our study were derived from highlighted. AETC – Adult emergency and Trauma Centre; Paediatric A&E – paediatric accident and emergency; PICU – paediatric intensive care unit; HDU – high dependency unit.
Fig 3. Most cases suspected to be caused by within-ward transmissions. Mapping trees against reference genomes for (**A**) *K. pneumoniae* subsp. *pneumoniae* excluding STs represented with only a single isolate in the dataset, (**B**) *K. quasipneumoniae* and (**C**) *K. variicola*. (**D-E**) show ST-specific mapping of the relevant isolates against references for ST35 (**D**), ST39 (**E**), ST14/ST15 (**F**). The coloured tips indicates isolates with <5 SNPs distance to their nearest neighbour based on a root-to-tip approach to identify related clusters of isolates(19). The colour ring in (**A**) shows the main STs, the columns in **D-F** show year of isolate (left column) and major ward (right column).
Fig 4. **Resistance mechanisms only explain beta-lactam, not aminoglycoside, phenotypes.** (**A**) The distribution of ESBL genes over time illustrating the dominance of $bla_{CTX\text{-}M\text{-}15}$. (**B**) Phenotypic chloramphenicol resistance profiles proportional over time and the presence or absence of *catA1* and/or *catA2* compared to *catB4*, the recently re-characterised non-functional CAT enzyme (Graf et al. 2023). (**C**) and (**D**) Predicted aminoglycoside resistance was based on the presence of an Aac6'-Ib-cr or Aac6'-Ib gene (AAC(6')-I) for amikacin and Aac3_IIa (AAC(3)-II) or AadB (ANT(2")-I) for gentamicin. The x-axis shows the

samples based on phenotypic resistance profiles (intermediate, sensitive, resistant) against (**C**) amikacin and (**D**) gentamicin, and the bars show number of isolates in the respective category that would have been predicted as sensitive (blue) or resistant (red) based on the genes as above.

Fig 5. High fluctuation of putative vaccine targets over time and diversity within major STs. The distribution of O-Ag locus types (top panels) and K-locus types (lower panels), resp., over time showing total numbers (**A**) and (**D**), only isolates with ESBL resistance genes (**B**) and (**E**) or ESBL-negative isolates (**C**) and (**F**).

S1 Fig. Age specific *Kpn* crude frequency and blood culture positivity rate. (**A**) Number of *Kpn* cases up to age 12 months (1998-2015). (**B**) Blood culture positivity rate (per 1,000 blood cultures) of *Kpn* cases up to age 12 months (1998-2015). (**C**) Number of *Kpn* cases up to age 12 months (2016-2021). (**D**) Blood culture positivity rate (per 1,000 blood cultures) of *Kpn* cases up to age 12 months (2016-2021). (**E**) Number of *Kpn* cases up to 100 years (1998-2015). (**F**) Blood culture positivity rate (per 1,000 blood cultures) of *Kpn* cases up to 100 years (1998-2015). (**G**) Number of *Kpn* cases up to 100 years (2016-2021). (**H**) Blood culture positivity rate (per 1,000 blood cultures) of *Kpn* cases up to 100 years (2016-2021).

S2 Fig. Quality control of sequencing data. The species assignment of the sequencing reads according to kraken are shown in panel (**A**), where all isolates with more than 10% of a different species were removed. (**B**) The plot shows the distribution of homozygous vs heterozygous SNPs, isolates represented by red points (>5% heterozygous SNPs) were removed as likely consisting of two related strains. (**C**) Number of contigs vs total assembly length is shown; with isolates that passed QC in black, isolates that failed the species composition QC in blue, and isolates with >500 contigs in green, which were also removed.

S3 Fig. Main ST isolates per wards over time. Each panel shows the number of isolates for the relevant ST as in the legend. The number of isolates per year are stratified by the major wards as in the figure legend.

S4 Fig. Phenotypic resistance profiles of all sequenced isolates. This highlights very high resistance against widely used first-line treatments like (**A**) cefpodoxime (ceftriaxone; 3GC), (**B**) augmentin (BLBLI combination), (**C**) gentamicin (AGly), (**D**) cotrimoxazole (TMT) but still a large proportion of sensitive isolates against less commonly used antimicrobials; (**E**) ciprofloxacin (Fq), (**F**) amikacin (AGly), and a pattern of increasing sensitivity against (**G**) chloramphenicol. The currently only available last-line treatments are (**H**) piperazillin-tazobactam (BLBLI combination), which show very high resistance levels; and (**I**) meropenem (carbapenem), the currently only viable alternative tested. (**J**) Cefoxitin also shows a lot of sensitive isolates; this antimicrobial is routinely used in diagnostic laboratories world-wide to test for *ampC* production as it was rapidly replaced with third-generation cephalosporins after its initial characterisation. The resistance profile is however of high interest as clinical trials have been started recently to assess it as a potential treatment alternative.

S5 Fig. Predicted resistance and virulence genes and plasmid replicons. The guidance tree was calculated with iqtree based on the core gene alignment derived from panaroo. The metadata columns displayed are *Kpn* species, main ST, sample type and year of isolation; as well as heatmaps showing the presence (dark shade) or absence (light shade) of (**A**) resistance genes, (**B**) virulence genes and (**C**) plasmid replicons.

S6 Fig. Plasmid replicons and resistance gene numbers. (**A**) *Kpn* isolates over time showing the number of plasmid replicons encoded, and (**B**) the number of isolates for numbers of resistance genes present over time.

S7 Fig. O-Ag serotypes over time and resistance profiles. The distribution of predicted O-Ag serotypes based on their genomic loci(30,31) over time for (**A**) all isolates, (**B**) non-ESBL and (**C**) ESBL⁻encoding isolates, and their distribution across (**D**) main STs and (**E**) K-locus types.

S8 Fig. Proportions of virulent isolates and resistant isolates over time have markedly different pattern. (**A**) The number of isolates with a kleborate-predicted virulence score > 1 (1 usually indicates yersiniabactin which is chromosomally fixed in several STs) which remains very low and stable over time and found sporadically across various wards, whereas (**B**) isolates with resistance scores 1 or over (indicating ESBL or, in the case of score of 2, carbapenemase present) rapidly increase over time and biased towards several wards. This is reflected in the proportions; whilst isolates encoding several virulence factors remain at very low proportion over time (**C**), we can see the rapid increase of resistant isolates which remains at a very high level since the initial increase (**D**).

Table S1: Accessions, metadata and kleborate predictions per strain.

Table S2: Antimicrobial gene predictions using ariba.

Table S3: Plasmid replicon predictions using ariba.

Table S4: Phenotypic resistance profiles.

Table S5: O- and K-antigen predictions using updated definitions in kaptive.

References

1. Liu L, Oza S, Hogan D, Chu Y, Perin J, Zhu J, et al. Global, regional, and national causes of under-5 mortality in 2000-15: an updated systematic analysis with implications for the Sustainable Development Goals. Lancet Lond Engl. 2016 Dec 17;388(10063):3027–35.

2. Sands K, Carvalho MJ, Portal E, Thomson K, Dyer C, Akpulu C, et al. Characterization of antimicrobial-resistant Gram-negative bacteria that cause neonatal sepsis in seven low- and middle-income countries. Nat Microbiol. 2021 Apr;6(4):512–23.

3. Musicha P, Cornick JE, Bar-Zeev N, French N, Masesa C, Denis B, et al. Trends in antimicrobial resistance in bloodstream infection isolates at a large urban hospital in Malawi (1998-2016): a surveillance study. Lancet Infect Dis. 2017;17(10):1042–52.

4. Salzberg NT, Sivalogan K, Bassat Q, Taylor AW, Adedini S, El Arifeen S, et al. Mortality Surveillance Methods to Identify and Characterize Deaths in Child Health and Mortality Prevention Surveillance Network Sites. Clin Infect Dis. 2019 Oct 9;69(Supplement_4):S262–73.

5. Wyres KL, Lam MMC, Holt KE. Population genomics of Klebsiella pneumoniae. Nat Rev Microbiol [Internet]. 2020 Feb 13 [cited 2020 Feb 24]; Available from: http://www.nature.com/articles/s41579-019-0315-1

6. Ellington MJ, Heinz E, Wailan AM, Dorman MJ, de Goffau M, Cain AK, et al. Contrasting patterns of longitudinal population dynamics and antimicrobial resistance mechanisms in two priority bacterial pathogens over 7 years in a single center. Genome Biol. 2019 02;20(1):184.

7. Lipworth S, Vihta KD, Chau K, Barker L, George S, Kavanagh J, et al. Ten-year longitudinal molecular epidemiology study of Escherichia coli and Klebsiella species bloodstream infections in Oxfordshire, UK. Genome Med. 2021 Sep 3;13(1):144.

8. Fostervold A, Hetland MAK, Bakksjø R, Bernhoff E, Holt KE, Samuelsen Ø, et al. A nationwide genomic study of clinical Klebsiella pneumoniae in Norway 2001-15: introduction and spread of ESBLs facilitated by clonal groups CG15 and CG307. J Antimicrob Chemother. 2022 Feb 23;77(3):665–74.

9. Thorpe H, Booton R, Kallonen T, Gibbon MJ, Couto N, Passet V, et al. One Health or Three? Transmission modelling of *Klebsiella* isolates reveals ecological barriers to transmission between humans, animals and the environment [Internet]. Microbiology; 2021 Aug [cited 2023 May 17]. Available from: http://biorxiv.org/lookup/doi/10.1101/2021.08.05.455249

10. Wyres KL, Nguyen TNT, Lam MMC, Judd LM, van Vinh Chau N, Dance DAB, et al. Genomic surveillance for hypervirulence and multi-drug resistance in invasive Klebsiella pneumoniae from South and Southeast Asia. Genome Med. 2020 Dec;12(1):11.

11. World Health Organization, editor. Antimicrobial resistance: global report on surveillance. Geneva, Switzerland: World Health Organization; 2014. 232 p.

12. Rodrigues C, Passet V, Rakotondrasoa A, Diallo TA, Criscuolo A, Brisse S. Description of Klebsiella africanensis sp. nov., Klebsiella variicola subsp. tropicalensis subsp. nov. and Klebsiella variicola subsp. variicola subsp. nov. Res Microbiol. 2019 May;170(3):165–70.

13. Gordon MA, Walsh AL, Chaponda M, Soko D, Mbvwinji M, Molyneux ME, et al. Bacteraemia and mortality among adult medical admissions in Malawi--predominance of non-typhi salmonellae and Streptococcus pneumoniae. J Infect. 2001 Jan;42(1):44–9.

14. Lester R, Musicha P, Kawaza K, Langton J, Mango J, Mangochi H, et al. Effect of resistance to third-generation cephalosporins on morbidity and mortality from bloodstream infections in Blantyre, Malawi: a prospective cohort study. Lancet Microbe. 2022 Dec;3(12):e922–30.

15. Cornick J, Musicha P, Peno C, Seager E, Iroh Tam PY, Bilima S, et al. Genomic investigation of a suspected Klebsiella pneumoniae outbreak in a neonatal care unit in sub-Saharan Africa. Microb Genomics. 2021 Nov;7(11).

16. Musicha P, Msefula CL, Mather AE, Chaguza C, Cain AK, Peno C, et al. Genomic analysis of Klebsiella pneumoniae isolates from Malawi reveals acquisition of multiple ESBL determinants across diverse lineages. J Antimicrob Chemother. 2019 May 1;74(5):1223–32.

17. Chung The H, Karkey A, Pham Thanh D, Boinett CJ, Cain AK, Ellington M, et al. A high-resolution genomic analysis of multidrug-resistant hospital outbreaks of Klebsiella pneumoniae. EMBO Mol Med. 2015 Mar;7(3):227–39.

18. Ejaz H, Wang N, Wilksch JJ, Page AJ, Cao H, Gujaran S, et al. Phylogenetic Analysis of Klebsiella pneumoniae from Hospitalized Children, Pakistan. Emerg Infect Dis. 2017 Nov;23(11):1872–5.

19. Wailan AM, Coll F, Heinz E, Tonkin-Hill G, Corander J, Feasey NA, et al. rPinecone: Define sub-lineages of a clonal expansion via a phylogenetic tree. Microb Genomics. 2019;5(4).

20. Gorrie CL, Da Silva AG, Ingle DJ, Higgs C, Seemann T, Stinear TP, et al. Key parameters for genomics-based real-time detection and tracking of multidrug-resistant bacteria: a systematic analysis. Lancet Microbe. 2021 Nov;2(11):e575–83.

21. Mangochi H, Tolhurst R, Simpson V, Kawaza K, Chidziwisano K, Feasey NA, et al. A qualitative study exploring hand hygiene practices in a neonatal unit in Blantyre, Malawi: implications for controlling healthcare-associated infections. Wellcome Open Res. 2022;7:146.

22. Sojo-Dorado J, López-Hernández I, Rosso-Fernandez C, Morales IM, Palacios-Baena ZR, Hernández-Torres A, et al. Effectiveness of Fosfomycin for the Treatment of Multidrug-Resistant *Escherichia coli* Bacteremic Urinary Tract Infections: A Randomized Clinical Trial. JAMA Netw Open. 2022 Jan 13;5(1):e2137277.

23. McGovern PC, Wible M, El-Tahtawy A, Biswas P, Meyer RD. All-cause mortality imbalance in the tigecycline phase 3 and 4 clinical trials. Int J Antimicrob Agents. 2013 May;41(5):463–7.

24. Veleba M, Schneiders T. Tigecycline Resistance Can Occur Independently of the *ramA* Gene in Klebsiella pneumoniae. Antimicrob Agents Chemother. 2012 Aug;56(8):4466–7.

25. Zhang R, Dong N, Huang Y, Zhou H, Xie M, Chan EWC, et al. Evolution of tigecycline- and colistin-resistant CRKP (carbapenem-resistant Klebsiella pneumoniae) in vivo and its persistence in the GI tract. Emerg Microbes Infect. 2018 Jul 9;7(1):127.

26. Roy S, Datta S, Viswanathan R, Singh AK, Basu S. Tigecycline susceptibility in Klebsiella pneumoniae and Escherichia coli causing neonatal septicaemia (2007-10) and role of an efflux pump in tigecycline non-susceptibility. J Antimicrob Chemother. 2013 May;68(5):1036–42.

27. He T, Wang R, Liu D, Walsh TR, Zhang R, Lv Y, et al. Emergence of plasmid-mediated high-level tigecycline resistance genes in animals and humans. Nat Microbiol. 2019 Sep;4(9):1450–6.

28. Huang Y, Rana AP, Wenzler E, Ozer EA, Krapp F, Bulitta JB, et al. Aminoglycoside-resistance gene signatures are predictive of aminoglycoside MICs for carbapenem-resistant Klebsiella pneumoniae. J Antimicrob Chemother. 2022 Feb 2;77(2):356–63.

29. Follador R, Heinz E, Wyres KL, Ellington MJ, Kowarik M, Holt KE, et al. The diversity of Klebsiella pneumoniae surface polysaccharides. Microb Genomics. 2016;2(8):e000073.

30. Kelly SD, Ovchinnikova OG, Müller F, Steffen M, Braun M, Sweeney RP, et al. Identification of a second glycoform of the clinically prevalent O1 antigen from Klebsiella pneumoniae. Proc Natl Acad Sci U S A. 2023 Jul 18;120(29):e2301302120.

31. Lam MMC, Wick RR, Judd LM, Holt KE, Wyres KL. Kaptive 2.0: updated capsule and lipopolysaccharide locus typing for the Klebsiella pneumoniae species complex. Microb Genomics [Internet]. 2022 Mar 21 [cited 2023 Jul 20];8(3). Available from: https://www.microbiologyresearch.org/content/journal/mgen/10.1099/mgen.0.000800

32. Orskov I, Fife-Asbury MA. New Klebsiella Capsular Antigen, K82, and the Deletion of Five of Those Previously Assigned. Int J Syst Bacteriol. 1977 Oct 1;27(4):386–7.

33. Edmunds PN. Further Klebsiella Capsule Types. J Infect Dis. 1954 Jan 1;94(1):65–71.

34. Edwards PR, Fife MA. Capsule Types of Klebsiella. J Infect Dis. 1952 Jul 1;91(1):92–104.

35. Lewis JM, Mphasa M, Banda R, Beale MA, Heinz E, Mallewa J, et al. Colonization dynamics of extended-spectrum beta-lactamase-producing Enterobacterales in the gut of Malawian adults. Nat Microbiol. 2022 Oct;7(10):1593–604.

36. Lewis JM, Mphasa M, Banda R, Beale MA, Mallewa J, Heinz E, et al. Genomic and antigenic diversity of colonizing Klebsiella pneumoniae isolates mirrors that of invasive isolates in Blantyre, Malawi. Microb Genomics. 2022 Mar;8(3):000778.

37. Feasey NA, Houston A, Mukaka M, Komrower D, Mwalukomo T, Tenthani L, et al. A Reduction in Adult Blood Stream Infection and Case Fatality at a Large African Hospital following Antiretroviral Therapy Roll-Out. Pett S, editor. PLoS ONE. 2014 Mar 18;9(3):e92226.

38. Lewis JM, Lester R, Mphasa M, Banda R, Edwards T, Thomson NR, et al. Emergence of carbapenemase-producing Enterobacteriaceae in Malawi. J Glob Antimicrob Resist. 2020 Mar;20:225–7.

39. Kumwenda GP, Sugawara Y, Abe R, Akeda Y, Kasambara W, Chizani K, et al. First Identification and genomic characterization of multidrug-resistant carbapenemase-producing Enterobacteriaceae clinical isolates in Malawi, Africa. J Med Microbiol. 2019 Dec 1;68(12):1707–15.

40. Zhang F, Li Q, Bai J, Ding M, Yan X, Wang G, et al. Heteroresistance to Amikacin in Carbapenem-Resistant Klebsiella pneumoniae Strains. Front Microbiol. 2021;12:682239.

41. Mankhomwa J, Tolhurst R, M'biya E, Chikowe I, Banda P, Mussa J, et al. A Qualitative Study of Antibiotic Use Practices in Intensive Small-Scale Farming in Urban and Peri-Urban Blantyre, Malawi: Implications for Antimicrobial Resistance. Front Vet Sci. 2022 May 24;9:876513.

42. Darlow CA, Farrington N, Johnson A, McEntee L, Unsworth J, Jimenez-Valverde A, et al. Flomoxef and fosfomycin in combination for the treatment of neonatal sepsis in the setting of highly prevalent antimicrobial resistance. J Antimicrob Chemother. 2022 Apr 27;77(5):1334–43.

43. Lipworth S, Vihta KD, Chau KK, Kavanagh J, Davies T, George S, et al. Ten Years of Population-Level Genomic *Escherichia coli* and *Klebsiella pneumoniae* Serotype Surveillance Informs Vaccine Development for Invasive Infections. Clin Infect Dis. 2021 Dec 16;73(12):2276–82.

44. Heinz E, Ejaz H, Bartholdson Scott J, Wang N, Gujaran S, Pickard D, et al. Resistance mechanisms and population structure of highly drug resistant Klebsiella in Pakistan during the introduction of the carbapenemase NDM-1. Sci Rep. 2019 Dec;9(1):2392.

45. Heinz E, Brindle R, Morgan-McCalla A, Peters K, Thomson NR. Caribbean multi-centre study of Klebsiella pneumoniae: whole-genome sequencing, antimicrobial resistance and virulence factors. Microb Genomics [Internet]. 2019 May 1 [cited 2020 Feb 24];5(5). Available from: https://www.microbiologyresearch.org/content/journal/mgen/10.1099/mgen.0.000266

46. Wantuch PL, Knoot CJ, Robinson LS, Vinogradov E, Scott NE, Harding CM, et al. Capsular polysaccharide inhibits vaccine-induced O-antigen antibody binding and function across both classical and hypervirulent K2:O1 strains of Klebsiella pneumoniae. Hakansson AP, editor. PLOS Pathog. 2023 May 5;19(5):e1011367.

47. Loraine J, Heinz E, De Sousa Almeida J, Milevskyy O, Voravuthikunchai SP, Srimanote P, et al. Complement Susceptibility in Relation to Genome Sequence of Recent Klebsiella pneumoniae Isolates from Thai Hospitals. mSphere. 2018 07;3(6).

48. Kol O, Wieruszeski JM, Strecker G, Fournet B, Zalisz R, Smets P. Structure of the O-specific polysaccharide chain of Klebsiella pneumoniae O1K2 (NCTC 5055) lipopolysaccharide. A complementary elucidation. Carbohydr Res. 1992 Dec 15;236:339–44.

49. Whitfield C, Richards JC, Perry MB, Clarke BR, MacLean LL. Expression of two structurally distinct D-galactan O antigens in the lipopolysaccharide of Klebsiella pneumoniae serotype O1. J Bacteriol. 1991 Feb;173(4):1420–31.

50. Kol O, Wieruszeski JM, Strecker G, Montreuil J, Fournet B, Zalisz R, et al. Structure of the O-specific polysaccharide chain from Klebsiella pneumoniae O1K2 (NCTC 5055) lipopolysaccharide. Carbohydr Res. 1991 Sep 18;217:117–25.

51. Wood DE, Salzberg SL. Kraken: ultrafast metagenomic sequence classification using exact alignments. Genome Biol. 2014 Mar 3;15(3):R46.

52. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol J Comput Mol Cell Biol. 2012 May;19(5):455–77.

53. Page AJ, De Silva N, Hunt M, Quail MA, Parkhill J, Harris SR, et al. Robust high-throughput prokaryote de novo assembly and improvement pipeline for Illumina data. Microb Genomics. 2016;2(8):e000083.

54. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinforma Oxf Engl. 2014 Jul 15;30(14):2068–9.

55. Hunt M, Mather AE, Sánchez-Busó L, Page AJ, Parkhill J, Keane JA, et al. ARIBA: rapid antimicrobial resistance genotyping directly from sequencing reads. Microb Genomics. 2017;3(10):e000131.

56. Inouye M, Dashnow H, Raven LA, Schultz MB, Pope BJ, Tomita T, et al. SRST2: Rapid genomic surveillance for public health and hospital microbiology labs. Genome Med. 2014;6(11):90.

57. Gupta SK, Padmanabhan BR, Diene SM, Lopez-Rojas R, Kempf M, Landraud L, et al. ARG-ANNOT, a new bioinformatic tool to discover antibiotic resistance genes in bacterial genomes. Antimicrob Agents Chemother. 2014;58(1):212–20.

58. McArthur AG, Waglechner N, Nizam F, Yan A, Azad MA, Baylay AJ, et al. The comprehensive antibiotic resistance database. Antimicrob Agents Chemother. 2013 Jul;57(7):3348–57.

59. Carattoli A, Zankari E, García-Fernández A, Voldby Larsen M, Lund O, Villa L, et al. In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. Antimicrob Agents Chemother. 2014 Jul;58(7):3895–903.

60. Lam MMC, Wick RR, Watts SC, Cerdeira LT, Wyres KL, Holt KE. A genomic surveillance framework and genotyping tool for Klebsiella pneumoniae and its related species complex. Nat Commun. 2021 Jul 7;12(1):4188.

61. Wick RR, Heinz E, Holt KE, Wyres KL. Kaptive Web: User-Friendly Capsule and Lipopolysaccharide Serotype Prediction for *Klebsiella* Genomes. Diekema DJ, editor. J Clin Microbiol. 2018 Apr 4;56(6):e00197-18, /jcm/56/6/e00197-18.atom.

62. Torsten Seemann. snippy: fast bacterial variant calling from NGS reads [Internet]. Available from: https://github.com/tseemann/snippy

63. Wu KM, Li LH, Yan JJ, Tsao N, Liao TL, Tsai HC, et al. Genome sequencing and comparative analysis of Klebsiella pneumoniae NTUH-K2044, a strain causing liver abscess and meningitis. J Bacteriol. 2009 Jul;191(14):4492–501.

64. Croucher NJ, Page AJ, Connor TR, Delaney AJ, Keane JA, Bentley SD, et al. Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. Nucleic Acids Res. 2015 Feb 18;43(3):e15–e15.

65. Price MN, Dehal PS, Arkin AP. FastTree: Computing Large Minimum Evolution Trees with Profiles instead of a Distance Matrix. Mol Biol Evol. 2009 Jul 1;26(7):1641–50.

66. Elliott AG, Ganesamoorthy D, Coin L, Cooper MA, Cao MD. Complete Genome Sequence of Klebsiella quasipneumoniae subsp. similipneumoniae Strain ATCC 700603. Genome Announc. 2016 May 26;4(3):e00438-16.

67. Pinto-Tomás AA, Anderson MA, Suen G, Stevenson DM, Chu FST, Cleland WW, et al. Symbiotic nitrogen fixation in the fungus gardens of leaf-cutter ants. Science. 2009 Nov 20;326(5956):1120–3.

68. Page AJ, Taylor B, Delaney AJ, Soares J, Seemann T, Keane JA, et al. SNP-sites: rapid efficient extraction of SNPs from multi-FASTA alignments. Microb Genomics. 2016;2(4):e000056.

69. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. Mol Biol Evol. 2015 Jan;32(1):268–74.

70. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods. 2017 Jun;14(6):587–9.

71. Pupko T, Pe'er I, Shamir R, Graur D. A fast algorithm for joint reconstruction of ancestral amino acid sequences. Mol Biol Evol. 2000 Jun;17(6):890–6.

72. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res. 2021 Jul 2;49(W1):W293–6.

73. Shen Z, Gao Q, Qin J, Liu Y, Li M. Emergence of an NDM-5-Producing Hypervirulent Klebsiella pneumoniae Sequence Type 35 Strain with Chromosomal Integration of an Integrative and Conjugative Element, ICE *Kp1*. Antimicrob Agents Chemother. 2019 Dec 20;64(1):e01675-19.

74. Cabanel N, Rosinski-Chupin I, Chiarelli A, Botin T, Tato M, Canton R, et al. Evolution of VIM-1-Producing Klebsiella pneumoniae Isolates from a Hospital Outbreak Reveals the Genetic Bases of the Loss of the Urease-Positive Identification Character. Beiko RG, editor. mSystems. 2021 Jun 29;6(3):e00244-21.

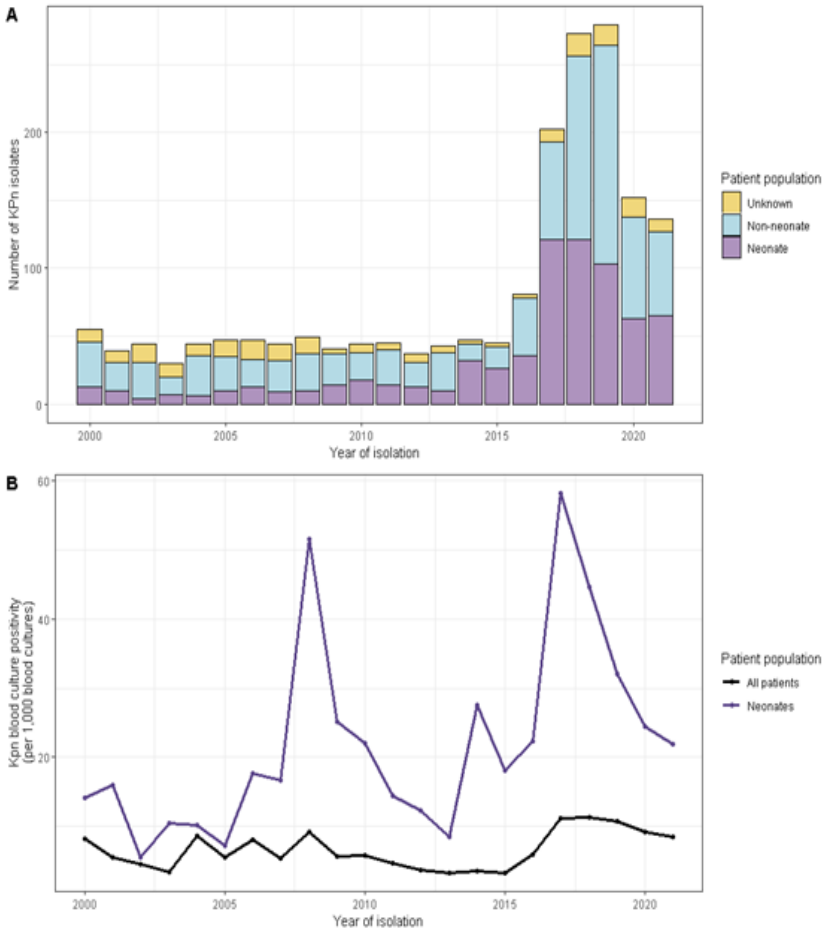
75. Stoesser N, Giess A, Batty EM, Sheppard AE, Walker AS, Wilson DJ, et al. Genome sequencing of an extended series of NDM-producing Klebsiella pneumoniae isolates from neonatal infections in a Nepali hospital characterizes the extent of community- versus hospital-associated transmission in an endemic setting. Antimicrob Agents Chemother. 2014 Dec;58(12):7347–57.

76. Tonkin-Hill G, MacAlasdair N, Ruis C, Weimann A, Horesh G, Lees JA, et al. Producing polished prokaryotic pangenomes with the Panaroo pipeline. Genome Biol. 2020 Dec;21(1):180.

77. Wickham H. Ggplot2: elegant graphics for data analysis. New York: Springer; 2009. 212 p. (Use R!).

78. Yu G, Lam TTY, Zhu H, Guan Y. Two Methods for Mapping and Visualizing Associated Data on Phylogeny Using Ggtree. Mol Biol Evol. 2018 01;35(12):3041–3.

79. R Core Team. A Language and Environment for Statistical Computing [Internet]. R Foundation for Statistical Computing, Vienna.; 2021. Available from: https://www.R-project.org


A
Number of KPn isolates
0
100
200
2000
2005
2010
2015
2020
Year of isolation
Patient population
Unknown
Non-neonate
Neonate
B
Kpn blood culture positivity (per 1,000 blood cultures)
20
40
60
2000
2005
2010
2015
2020
Year of isolation
Patient population
All patients
Neonates

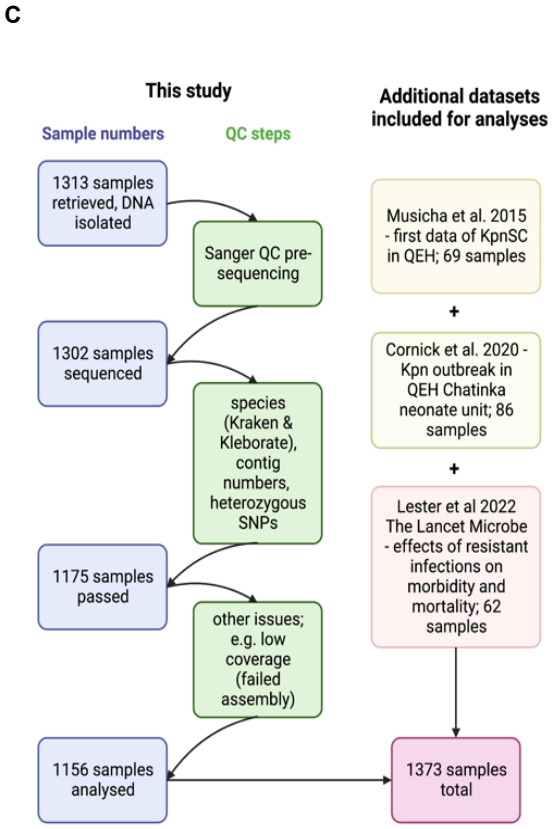
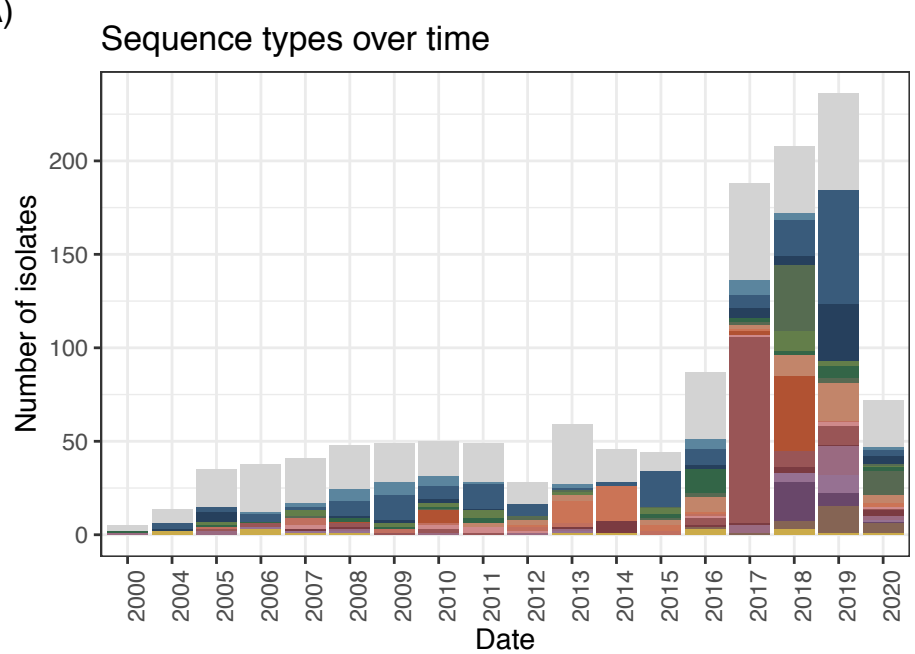
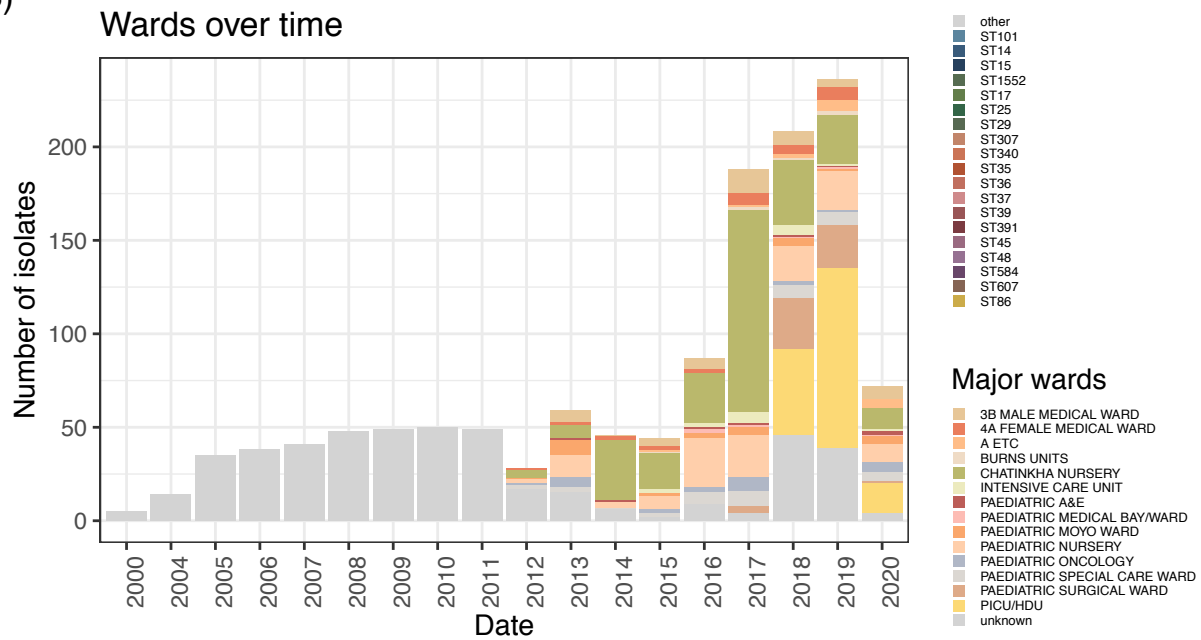

C
This study
Additional datasets included for analyses
Sample numbers
QC steps
1313 samples retrieved, DNA isolated
Sanger QC pre-sequencing
1302 samples sequenced
species (Kraken & Kleborate), contig numbers, heterozygous SNPs
1175 samples passed
other issues; e.g. low coverage (failed assembly)
1156 samples analysed
Musicha et al. 2015 - first data of KpnSC in QEH; 69 samples
+
Cornick et al. 2020 - Kpn outbreak in QEH Chatinka neonate unit; 86 samples
+
Lester et al 2022 The Lancet Microbe - effects of resistant infections on morbidity and mortality; 62 samples
1373 samples total

Figure 2

A)


Sequence types over time
Number of isolates
0
50
100
150
200
2000
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
Date

B)


Wards over time
Number of isolates
0
50
100
150
200
2000
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
Date
Main STs
other
ST101
ST14
ST15
ST1552
ST17
ST25
ST29
ST307
ST340
ST35
ST36
ST37
ST39
ST391
ST45
ST48
ST584
ST607
ST86
Major wards
3B MALE MEDICAL WARD
4A FEMALE MEDICAL WARD
A ETC
BURNS UNITS
CHATINKHA NURSERY
INTENSIVE CARE UNIT
PAEDIATRIC A&E
PAEDIATRIC MEDICAL BAY/WARD
PAEDIATRIC MOYO WARD
PAEDIATRIC NURSERY
PAEDIATRIC ONCOLOGY
PAEDIATRIC SPECIAL CARE WARD
PAEDIATRIC SURGICAL WARD
PICU/HDU
unknown

C)

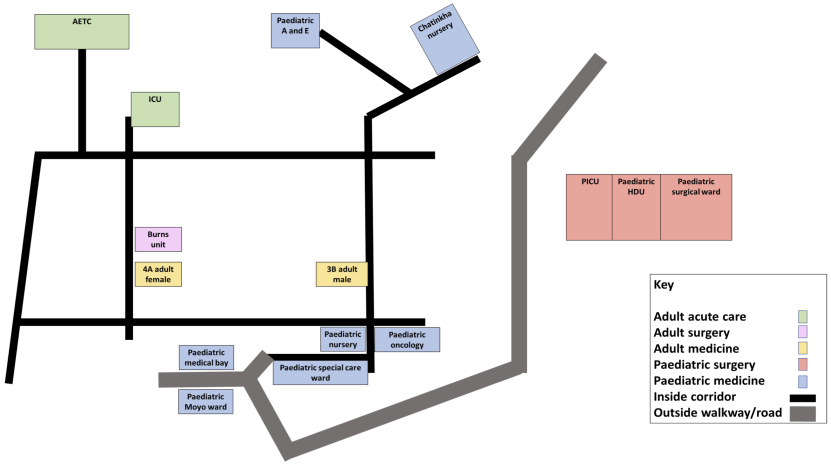
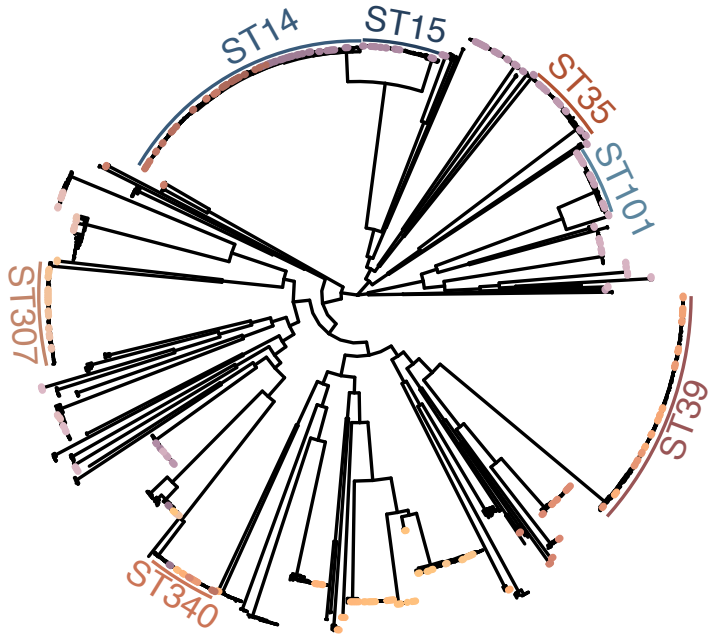
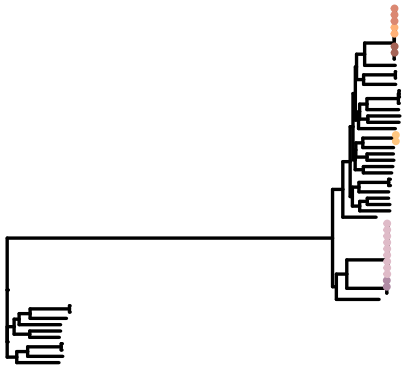

AETC
Paediatric A and E
Chatinkha nursery
ICU
PICU
Paediatric HDU
Paediatric surgical ward
Burns unit
4A adult female
3B adult male
Paediatric nursery
Paediatric oncology
Paediatric medical bay
Paediatric special care ward
Paediatric Moyo ward
Key
Adult acute care
Adult surgery
Adult medicine
Paediatric surgery
Paediatric medicine
Inside corridor
Outside walkway/road

Figure 3

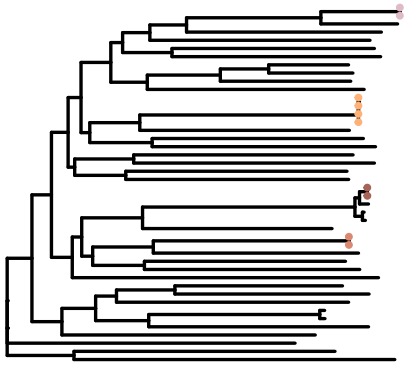
A) K. pneumoniae subsp. pneumoniae



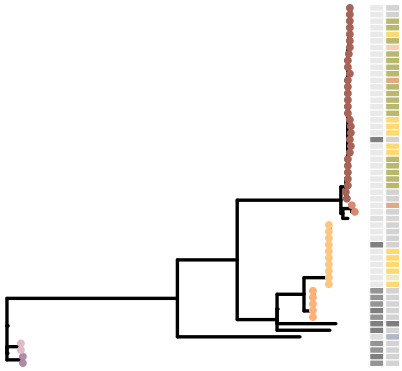
B) K. quasipneumoniae



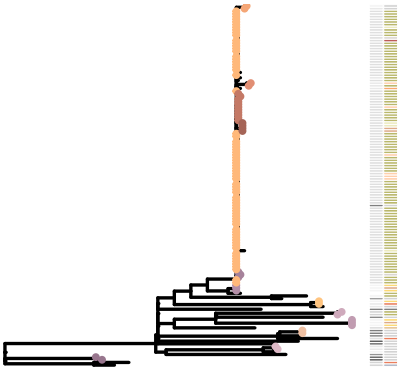
C) K. variicola



D) ST35



E) ST39



F) ST14 & ST15

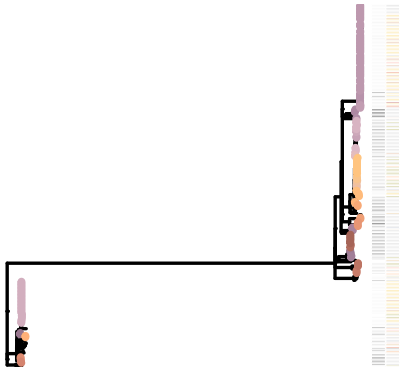


Figure 4

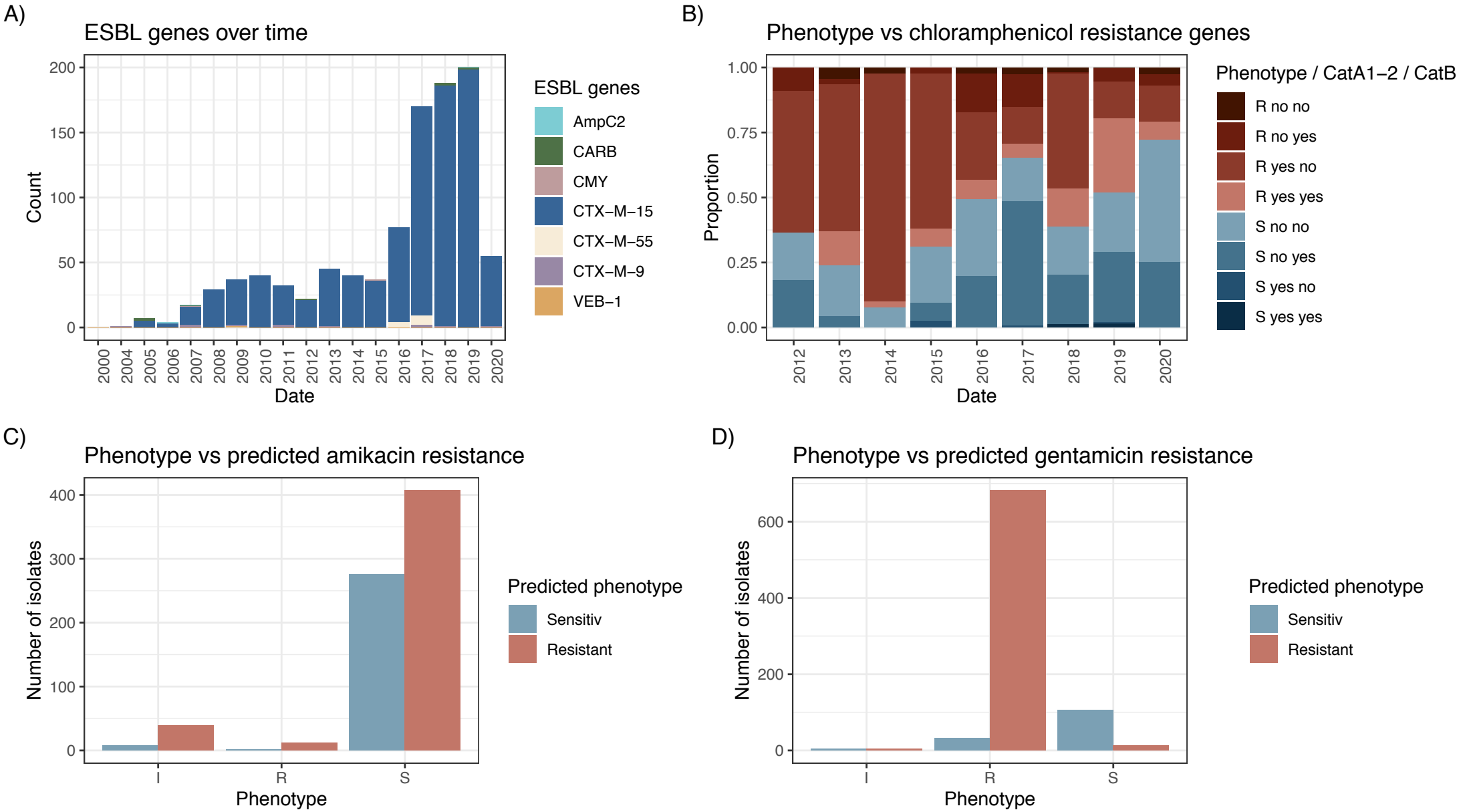
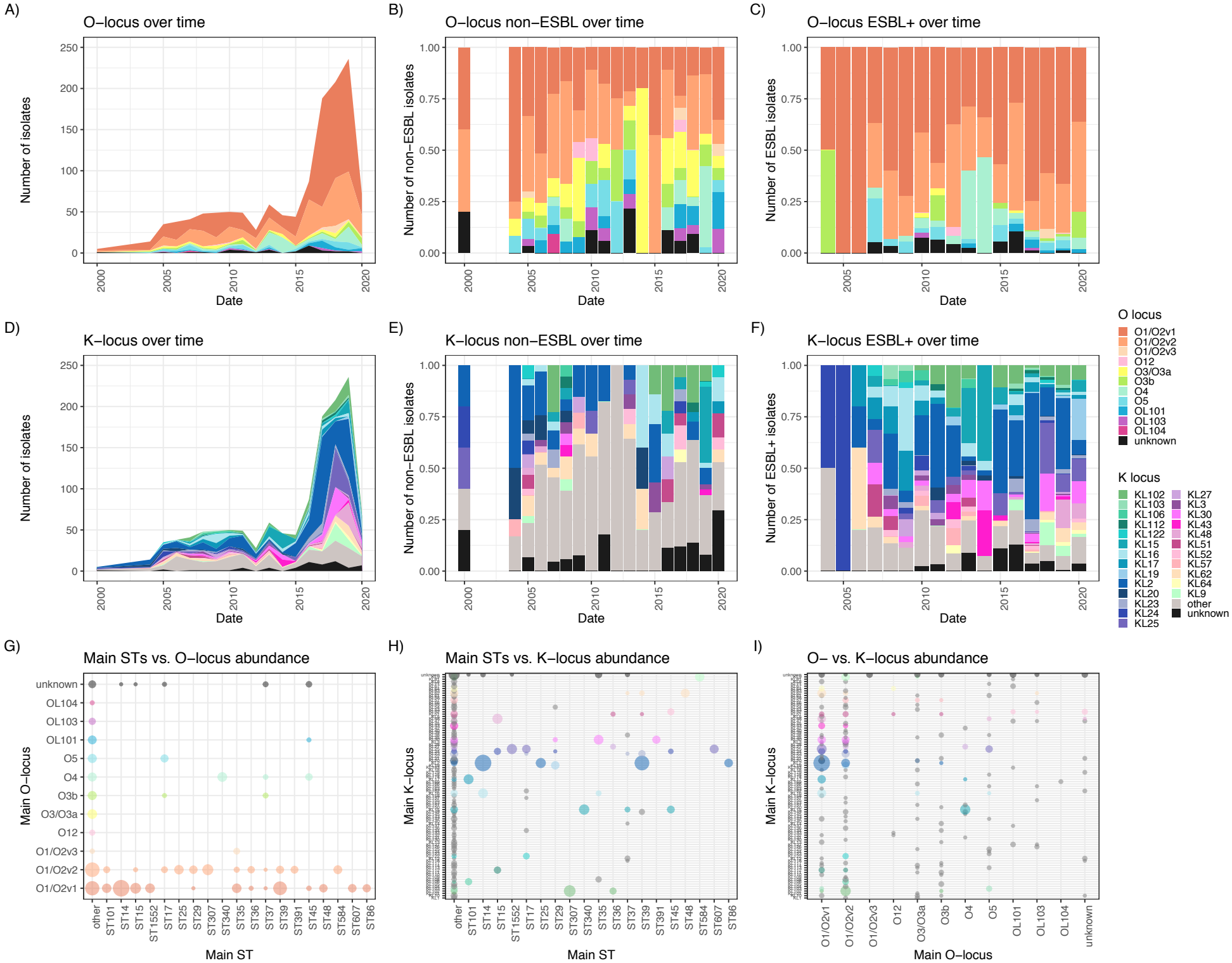

A)
ESBL genes over time
Count
Date
ESBL genes
AmpC2
CARB
CMY
CTX−M−15
CTX−M−55
CTX−M−9
VEB−1
B)
Phenotype vs chloramphenicol resistance genes
Proportion
Date
Phenotype / CatA1−2 / CatB
R no no
R no yes
R yes no
R yes yes
S no no
S no yes
S yes no
S yes yes
C)
Phenotype vs predicted amikacin resistance
Number of isolates
Phenotype
Predicted phenotype
Sensitiv
Resistant
D)
Phenotype vs predicted gentamicin resistance
Number of isolates
Phenotype
Predicted phenotype
Sensitiv
Resistant

Figure 5


A)
O−locus over time
Number of isolates
Date
B)
O−locus non−ESBL over time
Number of non−ESBL isolates
Date
C)
O−locus ESBL+ over time
Number of ESBL isolates
Date
D)
K−locus over time
Number of isolates
Date
E)
K−locus non−ESBL over time
Number of non−ESBL isolates
Date
F)
K−locus ESBL+ over time
Number of ESBL+ isolates
Date
O locus
O1/O2v1
O1/O2v2
O1/O2v3
O12
O3/O3a
O3b
O4
O5
OL101
OL103
OL104
unknown
K locus
KL102
KL103
KL106
KL112
KL122
KL15
KL16
KL17
KL19
KL2
KL20
KL23
KL24
KL25
KL27
KL3
KL30
KL43
KL48
KL51
KL52
KL57
KL62
KL64
KL9
other
unknown
G)
Main STs vs. O−locus abundance
Main O−locus
Main ST
H)
Main STs vs. K−locus abundance
Main K−locus
Main ST
I)
O− vs. K−locus abundance
Main K−locus
Main O−locus