Ten years of population-level genomic *Escherichia coli* and *Klebsiella pneumoniae* serotype surveillance informs vaccine development for invasive infections

Samuel Lipworth[1,2], Karina-Doris Vihta[1], Kevin K Chau[1], James Kavanagh[1], Timothy Davies[1,2], Sophie George[1], Leanne Barker[1], Ali Vaughan[1], Monique Andersson[2], Katie Jeffery[2], Sarah Oakley[2], Marcus Morgan[2], Timothy EA Peto[,2,3,4], Derrick W Crook[1,2,3,4], A Sarah Walker[1,3,4]*, Nicole Stoesser[1,2]*

* contributed equally

[1]Nuffield Department of Medicine, University of Oxford, Oxford, UK

[2]Oxford University Hospitals NHS Foundation Trust, Oxford, UK

[3]NIHR Health Protection Research Unit in Healthcare Associated Infections and Antimicrobial Resistance at University of Oxford in partnership with Public Health England, Oxford, United Kingdom

[4]NIHR Biomedical Research Centre, Oxford, United Kingdom

Corresponding Author: Dr Samuel Lipworth, samuel.lipworth@ndm.ox.ac.uk

Summary: A large, unselected, 10-year sequencing-based serotype analysis of *Escherichia coli* and *Klebsiella pneumoniae*-associated bloodstream infections (Oxfordshire, UK [n~3827] plus available global sequences [n=3678]). This strongly supports further development of effective O-antigen-targeted vaccines, which could substantially reduce BSIs and antimicrobial resistance.

Abstract

Background: The incidence of bloodstream infections (BSIs) caused by *Escherichia coli* and *Klebsiella pneumoniae* is increasing, with substantial associated morbidity, mortality and antimicrobial resistance. Unbiased serotyping studies to guide vaccine target selection are limited.

Methods: We conducted unselected, population-level genomic surveillance of bloodstream *E. coli* and *Klebsiella pneumoniae* isolates from 2008-2018 in Oxfordshire, UK. We supplemented this with an analysis of publicly available global sequencing data (n=3678).

Results: We sequenced 3478 *E. coli* isolates (3278 passed quality control) and 556 *K. pneumoniae* isolates (535 (K-antigen) and 549 (O-antigen) passed quality control). The four most common *E. coli* O-antigens (O1/O2/O6/O25) were identified in 1499/3278 isolates; the incidence of these O-types increased over time (IRRy=1.14, 95% CI:1.11-1.16). These O-types accounted for 616/1434 multidrug resistant (MDR) and 173/256 extended-spectrum beta-lactamase(ESBL)-resistant isolates in Oxfordshire, but only 19/90 carbapenem-resistant isolates across all studies. For *Klebsiella pneumoniae*, the most common O-antigens (O2v2/O1v1/O3b/O1v2) accounted for 410/549 isolates; the incidence of BSIs caused by these also increased annually (IRRy=1.09; 95% CI:1.05-1.12). These O-types accounted for 122/148 MDR and 106/123 ESBL isolates in Oxfordshire and 557/734 carbapenem-resistant isolates across all studies. Conversely

we observed substantial capsular antigen diversity. Analysis of 3678 isolates from

global studies demonstrated the generalisability of these findings. For *E. coli*, based on

serotyping, the ExPEC4V and ExPEC10V vaccines under investigation would cover 46%

and 72% of Oxfordshire isolates respectively, and 47% and 71% of MDR isolates.

Conclusions: O-antigen targeted vaccines may be useful in reducing the morbidity,

mortality and antimicrobial resistance associated with *E. coli* and *K. pneumoniae* BSIs.

Keywords: Vaccine; Bloodstream infection; Antimicrobial resistance; Whole genome

sequencing; Enterobacteriaceae

List of abbreviations:

CA - Community-associated

HA - Healthcare-associated

ST - Sequence Type

MDR - multidrug-resistant

ESBL - Extended Spectrum Beta Lactamase

BSI - Bloodstream Infection

E. coli - Escherichia coli

K. pneumoniae - Klebsiella pneumoniae

IORD - Infections in Oxfordshire Research Database

IRRy – Incidence rate ratio per year

Introduction

The incidence of bloodstream infections (BSIs) caused by *Enterobacteriaceae* (e.g. *Escherichia coli*, *Klebsiella pneumoniae*) continues to increase globally[1–3]. *Enterobacteriaceae* are a significant antimicrobial resistance (AMR) threat as they can acquire new genes horizontally and therefore rapidly adapt to changing selection pressures, including antimicrobial use. With the limited development of new antimicrobials to treat multidrug-resistant (MDR) infections, the threat of untreatable disease is a significant risk[4]. To date, a nationally mandated set of infection prevention and control standards arising from the Health and Social Care Act of 2008[5] has had limited success in reducing the incidence of BSIs caused by *Enterobacteriaceae*, as exemplified by national surveillance data[4,6].

Prophylactic vaccines represent an alternative approach to combating antimicrobial resistance (AMR) by reducing antibiotic usage and preventing infections caused by AMR-associated strains. For *E. coli*, the O-antigen, a component of lipopolysaccharide (LPS), has been considered the most promising vaccine target[7]. A recently developed bioconjugate vaccine (ExPEC4V; Janssen Pharmaceuticals) targeting four *E. coli* O-antigens (O1A, O2, O6A and O25B) has subsequently shown promise in a Phase II study[8]. A phase 1/2 clinical trial for a 10-valent vaccine (ExPEC10V) is currently recruiting[9]. For *K. pneumoniae,* vaccine development is lagging behind[10], and there are no vaccines currently in clinical trials.

For *E. coli*, recent serotyping studies have described limited O-antigen diversity in foodborne[11], ETEC/EPEC[11] and extended-spectrum beta-lactamase (ESBL)-producing BSI isolates[12]. To our knowledge however, there are no recent systematic and longitudinal studies characterising the sero-epidemiology of unselected BSIs. Recent studies of multi-country isolate collections which were non-systematically collected and/or enriched for antibiotic resistance have suggested that the O-antigen could also be a promising vaccine target for *K. pneumoniae*. Follador and colleagues analysed four sequencing datasets (one pre-1950, two others enriched for AMR/clonal outbreaks) and found that three O-antigens (O1/2/3) accounted for 80% of isolates whereas the diversity of K-antigens was an order of magnitude higher[13]. Similarly the O1/2/3/5 antigens accounted for 90.1% of isolates in a recent global sequencing study[14], though there were only relatively small numbers of isolates over limited timeframes from each country and it was unclear how these had been selected.

Large scale, systematic typing studies could help to inform the rational selection of vaccine targets across population groups which best mitigate the risk of BSI (including those associated with AMR) and provide a baseline for monitoring the emergence of serotype variation in response to vaccine roll-out. We performed a large, systematic study over ten years in Oxfordshire and combined this with an analysis of publicly available global datasets to provide a robust basis for future vaccine development.

Methods

Sample Collection/Sequencing

All blood culture isolates identified by the Oxford University Hospitals clinical microbiology laboratory between September 2008-December 2018 as being *K. pneumoniae/E. coli* (by matrix-assisted laser desorption/ionisation time-of-flight (MALDI-TOF)) analysis (from 2013) and by growth on chromogenic/MacConkey agar and analytical profile index tests (prior to 2013) were included in the study. One E. coli isolate was excluded due to a cataloging error. Infection control procedures were in place for inpatients known to be colonized with ESBL/carbapenemase-producing *Enterobacteriaceae* (CPE) during the study, including isolation in side rooms where feasible, enhanced contact precautions, and screening of high-risk patients on admission (for CPE). To avoid double counting instances where patients had several positive blood cultures taken during a single episode of infection, we considered only the first available isolate per 90 day period per patient.

Isolates were stored at -80°C in glycerol nutrient broth. Subcultures were grown overnight on Columbia blood agar (Oxoid, Basingstoke, UK) at 37°C. DNA for sequencing was extracted using the QuickGene DNA extraction kit (Autogen, MA, USA) as per the manufacturer's instructions with the addition of a mechanical lysis step (FastPrep, MP Biomedicals, CA, USA; 6m/s for 40 secs). Sequencing was performed using Illumina HiSeq 2500/3000/4000/MiSeq instruments as previously described[15]. To place our results in a global context, we additionally analysed 3678 isolates from existing studies selected to

encompass the recent national/international longitudinal datasets available in the literature as well as isolates that were relatively rare in our collection (e.g. carbapenemase producing/genotypically hypervirulent isolates; data available through NCBI project accession numbers: PRJNA480723[16], PRJEB4681[17], PRJEB35000[18], PRJEB12513[19], PRJEB30913[20], PRJEB1271[21], PRJEB10018[22,23]).

Statistics

Stacked negative binomial regression was used to compare incidence rate ratios over time (per year longer, IRRy) of BSI belonging to different antigen groups[24]. For each serotype or group of serotypes examined, we modelled the incidence of isolates with and without this trait over calendar time. In order to include as many isolates as possible in these models we defined calendar years as beginning in November (i.e. isolates before November 2008/after November 2018 are excluded from this part of the analysis). A Wald test for heterogeneity was performed to test the null hypothesis that the incidence trend did not differ between isolates with and without the trait. Healthcare-associated (HA) infection was defined as either nosocomial infection (onset >48 hours after admission) or onset within 30 days of last discharge and community-associated (CA) as that occurring >30 days since last admission.[25,26] Statistical analysis was performed using R version 4.0.2, except for stacked negative binomial regression models which were implemented in STATA version 15. Sequencing data was linked to electronic healthcare records via the Infections in Oxfordshire Research Database (IORD). IORD has generic Research Ethics Committee, Health Research Authority and Confidentiality Advisory Group approvals (14/SC/1069, ECC5-017(A)/2009).

Bioinformatics

De novo assembly was performed using Shovill (version 1.0.4)[27]. *In silico* serotyping was performed using the Ectyper tool[28–30] for *E. coli* and with Kleborate for *Klebsiella* spp.[31] For *E. coli* serotypes we included only those isolates which passed the ECtyper quality control (QC) filter and had a high confidence locus call; similarly for *Klebsiella* spp. we included only isolates where the Kleborate locus confidence call was at least 'good' (found in a single piece or with ≥95% coverage, with ≤ 3 missing genes and ≤ 1 extra genes)[31]. Genes (yersiniabactin, colibactin and aerobactin) associated with hypervirulence (defined as the ability to cause severe invasive disease in otherwise healthy individuals in the community[31]) were detected using the Kleborate tool. Virulence score was part of the Kleborate output, namely: 0 = no acquired virulence factors, 1= yersiniabactin only, 2 = yersiniabactin and colibactin (or colibactin only), 3 = aerobactin only, 4 = aerobactin and yersiniabactin, 5 = yersiniabactin, colibactin and aerobactin.

Antibiotic susceptibility testing

 Antibiotic susceptibility was measured using the BD Phoenix (Becton Dickinson, NJ USA) platform (EUCAST breakpoints[33]) for isolates after 2013 or disc diffusion (BSAC breakpoints[34]) for earlier isolates; all susceptibilities were measured using standard operating procedures in a UK Accreditation Service (UKAS) accredited laboratory.

Antibiotic susceptibility data were available through the Infections in Oxfordshire Research Database. For comparisons with existing studies, *in silico* antibiotic resistance gene detection (as called by Kleborate) was used instead of *in vitro* phenotyping because the latter results were not universally available.

Results

A total of 3479 *E. coli* isolates were sequenced of which 3278 passed *in silico* serotyping quality control. From these, 106 unique O-antigens were identified, though a relatively small number of these accounted for the majority of isolates (Fig.1/Table S1). The overall incidence of *E. coli* BSIs increased over the study period IRRy=1.14; 95% CI:1.11-1.16. The four most common serotypes (O1A, O2, O6A and O25B; i.e. those covered by the ExPEC4V vaccine) were identified in 1499/3278 (46%) isolates. Their incidence increased over time (IRRy=1.14; 95% CI:1.11-1.16); as did non-ExPEC4V serotypes (IRRy=1.13; 95% CI:1.10-1.16) ($p_{heterogeneity}$=0.7, Fig.S1). There was no evidence that O1/O2/O6/O25 serotypes differed in prevalence between CA and HA isolates (912/1969, 46% vs 576/1292, 45% respectively; p=0.3), or across five age-groups (<18 years 52/122 (43%), 18-39 years 98/222(44%) , 40-59 years 204/477 (43%), 60-79 years 577/1282 (45%), >80 years 557/1158(48%); p=0.3). The ExPEC10V serotypes (O1A, O2, O4, O6A, O8, O15, O16, O18A, O25B and O75) were identified in 2347/3278 (72%) isolates. This vaccine does not include O17 which was the 5th most common O-antigen type in our study, accounting for 181/3278 (6%) isolates.

Of the 3251 isolates for which in vitro antimicrobial sensitivity data was available, the ExPEC4V serotypes accounted for 191/331 (58%) of ESBL-producing and 394/834 (47%) of multi-drug resistant (MDR; resistance to ≥3 antibiotic classes) isolates. The ExPEC10V serotypes accounted for 592/834 (71%) of MDR and 255/331 (77%) of ESBL isolates. O25 was particularly associated with both ESBL (162/450, 36%) and MDR (249/450, 55%) phenotypes (Fig.2). Of the two carbapenem-resistant Oxford *E. coli* BSI isolates, neither had an O-antigen contained in the ExPEC4V/10V vaccines (one was O17 and the other O9).

Similarly we sequenced 556 *K. pneumoniae* (and *quasipneumoniae)* isolates; 93 unique K- and 14 unique O-loci were confidently identified in 535 (K-antigen)/549 (O-antigen) isolates passing quality control (Table S2/3). The incidence of *K. pneumoniae* BSIs also increased overall over the study period IRRy=1.09; 95%CI:1.05-1.12. To determine the theoretical feasibility of developing a *K. pneumoniae* vaccine and for comparison with *E. coli*, we again considered the top four K- and O-loci. KL2, KL22, KL24 and KL3 were identified in 136/535 (25%) isolates (Fig.1). There was no evidence that the incidence of BSIs caused by this group of capsular types changed over time (IRRy=1.03; 95%CI:0.94-1.13), nor that incidence trends varied between these and other *Klebsiella* spp. BSIs ($p_{heterogeneity}$=0.23). By contrast, O2v2, O1v1, O3b and O1v2 accounted for 410/549 (75%) isolates and the incidence of BSIs caused by these increased over time (IRRy=1.09; 95%CI:1.05-1.12), though again there was no evidence that these incidence trends differed from all other O-types ($p_{heterogeneity}$=0.8). A comparable Klebsiella vaccine to the ExPEC10V

(ie. composed of the ten most common O-antigens, O2v2, O1v1, O3b, O1v2, O2v1, O3/O3a, O4, O5, OL101, and one of O1/O2v2, O12, OL103 (which were tied for 10th most common O-antigen)), would cover 541/549 (99%) of isolates in our study.

KL3 and KL30 were the most common capsular antigens associated with the ESBL phenotype (43/123 [35%] and 12/123 [10%] respectively). KL3 was also the K-antigen most commonly associated with MDR (45/148 [30%]). In total the four most common K-antigens overall (KL2/242/24/3) accounted for 50/123 (41%) of ESBL and 65/148 (44%) of MDR isolates. In contrast the four most common O-antigens accounted for 106/123 (86%) of ESBL and 122/148 (82%) MDR isolates, with no evidence that the proportion of HA/CA BSI caused by these four antigens differed (179/238 [75%] vs 230/310 [74%], respectively; p=0.9).

Previous studies have suggested that clinically hypervirulent *K. pneumoniae* isolates (i.e. causing CA invasive infection) are particularly associated with the KL1/2 capsular types[35]. There were only two isolates with the KL1 capsular antigen in our study, only one of which carried the virulence-associated genes encoding for yersiniabactin, colibactin, aerobactin, salmochelin, and *rmpA*. Whilst KL2 did have the most isolates carrying hypervirulence genes of any capsular type, there were also 13/35 KL2 isolates with no hypervirulence loci (Fig.3). 36/37 KL1/2 isolates had O1v1 or O1v2 antigens (the other had O1/O2v2) meaning that a quadrivalent O-antigen-targeted vaccine would theoretically provide good coverage against this important group of isolates. Whilst there were more CA than HA

isolates for the KL2 capsular type (19/15), this was not the only K-type where this was true and the clinically hypervirulent phenotype was observed in 74 capsular types in total.

As the isolates from our study originated from a single region of the UK with a relatively low AMR burden, we compared our serotypes to those from recently published studies in the UK[17,19,21] and globally[16,18,20,22,23]. Overall, whilst there was some regional variation, findings supported the potential for O-antigen-targeted quadrivalent vaccines to contribute to substantial reductions in overall BSIs, as well as those caused by ESBL and MDR isolates, in both *E. coli* and *K. pneumoniae* (table 1). For *K. pneumoniae* they would likely provide a good level of protection against carbapenem-resistant and (genetically) hypervirulent isolates.

Conversely for *E. coli*, only 19/90 (21%) isolates carrying carbapenem resistance genes in all studies carried O-antigens covered by the ExPEC4V vaccine, with the O102 (n=18), O25 (n=11), O8 (n=11) and O160 (n=10) antigens predominating in these isolates. This was only slightly improved when considering the antigens in the ExPEC10V vaccine which would cover 35/90 (39%) isolates in all studies. Reflecting this, only 13/90 isolates carrying carbapenem resistance genes belonged to the major group of sequence types (STs) (131/95/73/69) which are responsible for most (predominantly CA) *E. coli* BSIs reported previously[16,17,19].

Discussion

A quadrivalent vaccine comprising the four most prevalent O-antigens for *E. coli* and *Klebsiella* spp. could theoretically have a significant impact on the incidence of BSIs and ESBL/MDR infections, as well as offering an alternative approach to help tackle the global epidemic of AMR in *Enterobacteriaceae*. Ten-valent vaccines are at a slightly earlier stage of development for *E. coli* but our data suggests they have the potential to make a significant impact on the incidence of invasive disease and associated antimicrobial resistance. Such a vaccine for *K. pneumoniae* would have covered nearly all isolates (541/549 [99%]) associated with invasive disease over the past decade in Oxfordshire.

Vaccines should also provide some protection against non-BSI infections which may evolve into BSIs, in particular urinary tract infections. As well as reducing morbidity and providing economic benefits, this could also be expected to lower antibiotic selection pressures against *E. coli* and *Klebsiella* spp.. The relatively much higher incidence of these primary infections vs BSIs may increase the economic viability of any prospective vaccine which, particularly for *Klebsiella* spp., may be a significant barrier to development given the difficulty in identifying groups of patients in which vaccination would be cost-effective[36].

We would expect the ExPEC4V/ExPEC10V vaccines to offer less protection against carbapenem-resistant *E. coli* infections. Whilst at first glance this looks disappointing, the prevention of ESBL infections would still reduce the carbapenem selection pressure.

Additionally, it was notable that the majority of carbapenem resistance gene carrying *E. coli* isolates were not from the sequence types which have been found to cause most (predominantly CA) BSIs in other studies[16,17,19]. This might indicate that these isolates do not originate from the community reservoir and that they may be associated with nosocomial transmission as was found to be the case in *K. pneumoniae*[23]. Widespread dissemination of carbapenem resistance in sequence types associated with global epidemics of CA-MDR disease (eg ST131) would be a public health challenge and a vaccine may have a role in preventing this.

The main limitation of our study is that it is from a single region. To mitigate this we analysed data from other published studies; however, in contrast to our long-term regional surveillance most of these were either deliberately enriched for MDR isolates or at high risk of bias due to unclear sampling methodologies. Whilst direct comparison with our data is therefore difficult, the overall findings appear to be reflected both nationally and globally. We were unable to characterise the distribution of capsular (K) antigen types for *E. coli* because to our knowledge there is no curated database of sequences currently available. Additionally we were unable to assign an in silico O-antigen call to ~5% of isolates in both *E. coli* and *K. pneumoniae*. This may be due to assembly errors or the presence of novel serotypes.

Overall, O-antigens targeted vaccines for *E. coli* and *Klebsiella* spp. could theoretically have a significant impact on the incidence of BSIs and ESBL/MDR infections, as well as offering an alternative approach to help tackle the global epidemic of AMR in *Enterobacteriaceae*. Our findings support future development and subsequent efficacy studies for such vaccines, provided they are demonstrated to be safe and immunogenic. The substantial capsular antigen diversity observed for *Klebsiella* spp. BSIs would likely preclude this as a useful vaccine target. Surveillance following the implementation of any vaccine would be essential in identifying any evidence of strain replacement.

Authors' Notes

Competing interests: The authors declare that there are no competing interests.

References

1   Buetti N, Atkinson A, Marschall J, Kronenberg A, the Swiss Centre for Antibiotic Resistance (ANRESIS). Incidence of bloodstream infections: a nationwide surveillance of acute care hospitals in Switzerland 2008–2014. *BMJ Open* 2017; **7**: e013665.

2   Diekema DJ, Hsueh P-R, Mendes RE, *et al.* The Microbiology of Bloodstream Infection: 20-Year Trends from the SENTRY Antimicrobial Surveillance Program. *Antimicrob Agents Chemother* 2019; **63**. DOI:10.1128/AAC.00355-19.

3   Douglas NM, Hennessy JN, Currie BJ, Baird RW. Trends in Bacteremia Over 2 Decades in the Top End of the Northern Territory of Australia. *Open Forum Infect Dis* 2020; **7**: ofaa472.

4   English Surveillance Programme for Antimicrobial Utilisation and Resistance (ESPAUR). Public Health England, 2019 https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attac hment_data/file/843129/English_Surveillance_Programme_for_Antimicrobial_Utilisati on_and_Resistance_2019.pdf.

5   DHSC. The Health and Social Care Act 2008 Code of Practice of the prevention and control of infections and related guidance. Department of Health and Social Care, UK Goverment, 2015.

6   Preventing healthcare associated Gram-negative bloodstream infections: an improvement resource. Public Health England, 2017 https://www.england.nhs.uk/wp-content/uploads/2020/08/Gram-negative_IPCresource_pack.pdf.

7   Poolman JT, Wacker M. Extraintestinal Pathogenic Escherichia coli, a Common Human Pathogen: Challenges for Vaccine Development and Progress in the Field. *J Infect Dis* 2016; **213**: 6–13.

8   Frenck RW Jr, Ervin J, Chu L, *et al.* Safety and immunogenicity of a vaccine for extra-intestinal pathogenic Escherichia coli (ESTELLA): a phase 2 randomised controlled trial. *Lancet Infect Dis* 2019; **19**: 631–40.

9   A Study of Three Different Doses of VAC52416 (ExPEC10V) in Adults Aged 60 to 85 Years in Stable Health. clinicaltrials.gov. https://clinicaltrials.gov/ct2/show/NCT03819049 (accessed July 17, 2020).

10   Choi M, Tennant SM, Simon R, Cross AS. Progress towards the development of Klebsiella vaccines. *Expert Rev Vaccines* 2019; **18**: 681–91.

11   Ingle DJ, Valcanis M, Kuzevski A, *et al.* In silico serotyping of E. coli from short read data identifies limited novel O-loci but extensive diversity of O:H serotype combinations within and between pathogenic lineages. *Microb Genom* 2016; **2**: e000064.

12   Mamani R, Flament-Simon SC, García V, *et al.* Sequence Types, Clonotypes, Serotypes, and Virotypes of Extended-Spectrum β-Lactamase-Producing Escherichia coli Causing Bacteraemia in a Spanish Hospital Over a 12-Year Period (2000 to 2011). *Front Microbiol* 2019; **10**: 1530.

13   Follador R, Heinz E, Wyres KL, *et al.* The diversity of Klebsiella pneumoniae surface polysaccharides. *Microb Genom* 2016; **2**: e000073.

14   Choi M, Hegerle N, Nkeze J, *et al.* The Diversity of Lipopolysaccharide (O) and Capsular Polysaccharide (K) Antigens of Invasive Klebsiella pneumoniae in a Multi-Country Collection. *Front Microbiol* 2020; **11**: 1249.

15   De Maio N, Shaw LP, Hubbard A, *et al.* Comparison of long-read sequencing technologies in the hybrid assembly of complex bacterial genomes. *Microb Genom* 2019; **5**. DOI:10.1099/mgen.0.000294.

16   Hastak P, Cummins ML, Gottlieb T, *et al.* Genomic profiling of Escherichia coli isolates from bacteraemia patients: a 3-year cohort study of isolates collected at a Sydney teaching hospital. *Microb Genom* 2020; **6**. DOI:10.1099/mgen.0.000371.

17   Kallonen T, Brodrick HJ, Harris SR, *et al.* Systematic longitudinal survey of invasive Escherichia coli in England demonstrates a stable population structure only transiently disturbed by the emergence of ST131. *Genome Res* 2017; published online July 18. DOI:10.1101/gr.216606.116.

18   van Hout D, Verschuuren TD, Bruijning-Verhagen PCJ, *et al.* Extended-spectrum beta-lactamase (ESBL)-producing and non-ESBL-producing Escherichia coli isolates causing bacteremia in the Netherlands (2014 - 2016) differ in clonal distribution, antimicrobial resistance gene and virulence gene content. *PLoS One* 2020; **15**: e0227604.

19   Goswami C, Fox S, Holden M, Connor M, Leanord A, Evans TJ. Genetic analysis of invasive Escherichia coli in Scotland reveals determinants of healthcare-associated versus community-acquired infections. *Microb Genom* 2018; **4**. DOI:10.1099/mgen.0.000190.

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

21 Ellington MJ, Heinz E, Wailan AM, *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; **20**: 184.

22 Grundmann H, Glasner C, Albiger B, *et al.* Occurrence of carbapenemase-producing Klebsiella pneumoniae and Escherichia coli in the European survey of carbapenemase-producing Enterobacteriaceae (EuSCAPE): a prospective, multinational study. *Lancet Infect Dis* 2017; **17**: 153–63.

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

24 Lunn M, McNeil D. Applying Cox regression to competing risks. *Biometrics* 1995; **51**: 524–32.

25 Cardoso T, Almeida M, Friedman ND, *et al.* Classification of healthcare-associated infection: a systematic review 10 years after the first proposal. *BMC Med* 2014; **12**: 40.

26 Vihta K-D, Stoesser N, Llewelyn MJ, *et al.* Trends over time in Escherichia coli bloodstream infections, urinary tract infections, and antibiotic susceptibilities in Oxfordshire, UK, 1998–2016: a study of electronic health records. *Lancet Infect Dis* 2018; **18**: 1138–49.

27 Seemann T. Shovill. https://github.com/tseemann/shovill (accessed May 18, 2020).

28 Inouye M, Dashnow H, Raven L-A, *et al.* SRST2: Rapid genomic surveillance for public health and hospital microbiology labs. *Genome Med* 2014; **6**: 90.

29 Joensen KG, Tetzschner AMM, Iguchi A, Aarestrup FM, Scheutz F. Rapid and Easy In Silico Serotyping of Escherichia coli Isolates by Use of Whole-Genome Sequencing Data. *J Clin Microbiol* 2015; **53**: 2410–26.

30 ECTyper. https://github.com/phac-nml/ecoli_serotyping/tree/master/ectyper (accessed May 27, 2020).

31 Holt K. Kleborate. https://github.com/katholt/Kleborate (accessed May 22, 2020).

32 Wyres KL, Lam MMC, Holt KE. Population genomics of Klebsiella pneumoniae. *Nat Rev Microbiol* 2020; **18**: 344–59.

33   EUCAST. European Committee on Antimicrobial Susceptibility Testing.
      https://www.eucast.org/fileadmin/src/media/PDFs/EUCAST_files/Breakpoint_tables
      /v_10.0_Breakpoint_Tables.pdf.

34   Andrews JM, Howe RA, BSAC Working Party on Susceptibility Testing. BSAC
      standardized disc susceptibility testing method (version 10). *J Antimicrob Chemother*
      2011; **66**: 2726–57.

35   Choby JE, Howard-Anderson J, Weiss DS. Hypervirulent Klebsiella pneumoniae -
      clinical and molecular perspectives. *J Intern Med* 2020; **287**: 283–300.

36   Boston Consulting Group and The Wellcome Trust. Vaccines to tackle drug resistant
      infections An evaluation of R&D opportunities. https://vaccinesforamr.org/wp-
      content/uploads/2018/09/Vaccines_for_AMR.pdf.

FIGURE LEGENDS

Figure 1: Limited O-antigen diversity in Oxfordshire BSIs with broadly stable population structure over time. Proportion of A. *E. coli* O-antigens B. *Klebsiella* spp. capsular antigens and C. *Klebsiella* spp. O-antigens observed in all isolates over the ten year period. In D.-F., The red/green dashed lines denote the x-axis position of the four/ten most prevalent O/K-types and the blue dashed lines demonstrate the cumulative percentage of isolates which have these antigens, for *E. coli* O-antigens (D.), *Klebsiella* spp. capsular (K) antigens (E.), and *Klebsiella* spp. O-antigens (F.)

Figure 2: Left: number of fully-sensitive (to antibiotics shown in the right-hand upset plot) *E. coli* BSI isolates by O-type, Right: Upset plot showing the number of *E. coli* isolates with phenotypic resistance to the antibiotics shown, stratified by O-antigen type (colors) and ordered by number of intersections. The horizontal bar plot shows the total number of isolates phenotypically resistance to each antibiotic in any combination.

Figure 3: Left: number of *K. pneumoniae* BSI isolates either fully susceptible or resistant only to amoxicillin shown by O-antigen type. The above violin/jitter plot above shows the distribution of virulence scores, points are colored according to their O-antigen type. Right: Upset plot showing the number of *K. pneumoniae* isolates with resistance to the classes of antibiotics shown. The stacked bar plot shows the distribution of these by O-antigen type and the violin plot (with jittered points to show individual data points) above this demonstrates the corresponding distributions of virulence scores. Virulence score is as defined by Kleborate (see methods).

E. coli - O antigen	Proportion from O1/O2/O6/O25						
Study	**Oxford**	**CUH**[10,14]	**EUSCAPE**[15,16]	**SE Asia**[13]	**Netherlands**[11]	**Australia**[9]	**Scotland**[12]
Isolate group							
All	1499/3278 (46%)	205/399 (51%)	130/329 (40%)	NA	188/445 (42%)	44/79 (56%)	64/159 (40%)
ESBL	173/256 (68%)	30/36 (83%)	61/154 (40%)	NA	113/244 (46%)	12/15 (80%)	12/15 (80%)
MDR	616/1434 (43%)	95/192 (49%)	87/244 (36%)	NA	113/274 (41%)	26/44 (59%)	27/85 (32%)
Carbapenem resistant	0/2 (0%)	NA	18/86 (21%)	NA	1/2 (50%)	NA	NA
K. pneumoniae - K antigen	Proportion from KL2/KL242/KL24/KL3						

All	136/535 (25%)	21/152 (14%)	261/1634 (16%)	56/243 (23%)	NA	NA	NA
ESBL	50/123 (41%)	8/94 (9%)	124/856 (14%)	30/107 (28%)	NA	NA	NA
MDR	65/148 (44%)	12/110 (11%)	181/1221 (15%)	37/130 (28%)	NA	NA	NA
Carbapenem resistant	1/3 (33%)	0/2 (0%)	81/678 (12%)	2/15 (13%)	NA	NA	NA
Hypervirulent	15/28 (54%)	0/1 (0%)	21/55 (38%)	20/72 (28%)	NA	NA	NA
K. pneumoniae - O antigen	Proportion from O2v2, O1v1, O3b and O1v2						
All	410/549 (75%)	119/154 (77%)	1173/1651 (71%)	228/318 (72%)	NA	NA	NA
ESBL	106/123 (86%)	84/96 (88%)	592/864 (68%)	116/157 (74%)	NA	NA	NA
MDR	122/148 (82%)	98/112 (88%)	884/1231 (72%)	137/191 (72%)	NA	NA	NA

Carbapenem resistant	0/3 (0%)	2/2 (100%)	521/684 (76%)	34/45 (76%)	NA	NA	NA
Hypervirulent	27/28 (96%)	1/1 (100%)	47/55 (85%)	70/87 (80%)	NA	NA	NA

Table 1. Comparison of the Oxford BSIs with those from similar national/regional studies. For each group we considered the proportion of isolates in each study carrying the four most prevalent antigens in the Oxford dataset to consider the hypothetical coverage a vaccine targeted at these might provide. CUH - Cambridge University Hospitals, EUSCAPE - The European Survey of Carbapenemase-Producing Enterobacteriaceae, SE Asia - South-east Asia. 3479 *E. coli* isolates from Oxford were sequenced of which 3278 passed Ectyper QC. 556 Oxford *K. pneumoniae* isolates were sequenced of which 535 (K-antigen)/549 (O-antigen) passed Kleborate QC. 1510 E. coli isolates from published datasets were included of which 1411 passed Ectyper QC. Of the 2168 *K. pneumoniae* isolates included from published datasets, 2029 (K-antigen)/2123 (O-antigen) passed Klerorate QC typing. The ESBL/MDR/Carbapenem resistant numbers are based on gene detection from Kleborate, including for Oxford isolates (in contrast to the main text where we report the results of *in vitro* drug susceptibility testing).