

1 Molecular epidemiology of *Escherichia coli* and *Klebsiella*
2 species bloodstream infections in Oxfordshire (UK) 2008-
3 2018.

4 **Keywords:** Gram-negative bloodstream infections; Bacteraemia; Whole genome sequencing,
5 *Klebsiella pneumoniae*; virulence; antimicrobial resistance

6
7 Samuel Lipworth^{1,2}, Karina-Doris Vihta¹, Kevin Chau¹, Leanne Barker¹, Sophie George¹, James
8 Kavanagh¹, Timothy Davies^{1,2}, Alison Vaughan¹, Monique Andersson², Katie Jeffery², Sarah
9 Oakley², Marcus Morgan², Susan Hopkins⁴, Timothy E. A. Peto^{1,2,3,5}, Derrick W. Crook^{1,2,3,5}, Ann
10 Sarah Walker^{*1,3,5} and Nicole Stoesser^{*1,2}

11
12 *Joint senior authors

13
14 **Affiliations:**

15 ¹Nuffield Department of Medicine, University of Oxford, Oxford, United Kingdom

16 ²Oxford University Hospitals NHS Foundation Trust, Oxford, United Kingdom

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

20 ⁴National Infection Service, Public Health England, Colindale, London

21 ⁵NIHR Biomedical Research Centre, Oxford, United Kingdom

22

1

23

24 Abstract

25 The incidence of Gram-negative bloodstream infections (BSIs), predominantly caused by
26 *Escherichia coli* and *Klebsiella* species, continues to increase; however the causes of this are
27 unclear and effective interventions are therefore hard to design. In this study we sequenced
28 3468 sequential, unselected isolates over a decade in Oxfordshire, UK. We demonstrate that
29 the observed increases in *E. coli* incidence were not driven by clonal expansion; instead, four
30 major sequence types (STs) continue to dominate a stable population structure, with no
31 evidence of adaptation to hospital/community settings. Conversely in *Klebsiella* spp. most
32 infections are caused by sporadic STs with the exception of a local drug-resistant outbreak
33 strain (ST490). Virulence elements are highly structured by ST in *E. coli* but not *Klebsiella* spp.
34 where they occur in a diverse spectrum of STs and equally across healthcare and community
35 settings. Most clinically hypervirulent (i.e. community-onset) *Klebsiella* BSIs have no known
36 acquired virulence loci. Finally we demonstrate a diverse but largely genus-restricted mobilome
37 with close associations between antimicrobial resistance (AMR) genes and insertion sequences
38 but not typically specific plasmid replicon types; consistent with the dissemination of AMR genes
39 being highly contingent on smaller mobile genetic elements (MGEs). Our large genomic study
40 highlights distinct differences in the molecular epidemiology of *E. coli* and *Klebsiella* BSIs, and
41 suggests that no single specific pathogen genetic factors are likely contributing to the increasing
42 incidence of BSI overall, that association with AMR genes in *E. coli* is a contributor to the
43 increasing number of *E. coli* BSIs, and that more attention should be given to AMR gene
44 associations with non-plasmid MGEs to try and understand horizontal gene transfer networks.

45

46 Background

47 Gram-negative bloodstream infections (GNBSI), predominantly caused by *Escherichia coli* and
48 *Klebsiella* spp., are a significant and increasing threat to public health. They are now the leading
49 cause of bloodstream infection (BSI) in the UK with a substantial associated burden of morbidity
50 and mortality^{1,2}. Despite becoming a significant public health concern and policy focus, their
51 incidence continues to increase and it is unclear how ambitious targets to reduce this can be
52 achieved³.

53

54 There is a significant association between GNBSIs and antimicrobial resistance (AMR),
55 particularly in certain globally successful sequence types (STs), such as ST131^{4,5}.

56 Enterobacteriaceae have relatively open pan-genomes and are able to rapidly adapt to
57 changing selection pressures (including antibiotic usage)^{6,7}. Multidrug resistance (MDR)-
58 associated STs have been linked with prolonged hospital stay and adverse outcomes⁸. Whilst
59 infections caused by relatively susceptible isolates still represent the majority of cases, the
60 potential for rapid proliferation of AMR-associated clones and the dissemination of AMR genes
61 on mobile genetic elements between lineages and species is a major concern⁹.

62 Recent molecular epidemiology studies in the UK have replicated global findings that most *E.*
63 *coli* BSIs are caused by the clonal lineages ST131, 95, 73 (all phylogroup B2) and ST 69
64 (phylogroup D)^{7,10}. One study has shown that after the emergence and clonal expansion of
65 ST69 and ST131 in the early 2000s, the population structure reached an equilibrium, with STs
66 131, 95, 73 and 69 predominating¹¹. However, this study only sequenced isolates cultured prior
67 to 2012, which may represent a critical inflection point in the incidence rate,¹ which continues to

68 increase year on year. It remains unclear whether this is caused by homogenous expansion of
69 *E. coli* lineages or clonal expansion of an existing or emergent ST.

70
71 For *Klebsiella* spp., it has been hypothesised that isolates broadly cause two categories of
72 BSI^{12,13}, namely: (i) multidrug resistant healthcare-associated (HA) infections caused by low
73 virulence strains with AMR gene-associated plasmids, and (ii) clinically hypervirulent
74 community-associated (CA) infections caused by strains carrying virulence gene plasmids. The
75 convergence of these two phenotypes might pose a significant risk to human health¹⁴. Whilst
76 there are multiple studies describing in detail the epidemiology of globally distributed clones
77 associated with AMR and hyper-virulence^{5,15}, such isolates represent a minority of BSI in the UK
78 and most of Europe¹⁶. A recent study in Cambridgeshire (UK) of 162 *K. pneumoniae* isolates,
79 enriched to over-represent MDR isolates, demonstrated a predominance of globally important
80 clones with apparent cycling of relative incidence over a two-year period¹⁷. However, MDR
81 infections represent approximately 18% of *K. pneumoniae* BSIs in England and the molecular
82 epidemiology of most invasive disease caused by *Klebsiella* spp. has not been systematically
83 studied².

84
85 In order to investigate underlying potential pathogen genetic factors driving the increasing
86 incidence of these infections, we analysed all *Klebsiella* spp. and *E. coli* BSIs over a decade
87 (2008-2018) in Oxfordshire, UK, linking electronic health records and mandatory reporting data
88 with sequencing data for all isolates, to give a detailed picture of the regional genomic
89 epidemiology of *E. coli* and *Klebsiella* spp. BSIs. We used other publicly available sequencing
90 datasets to contextualise our findings.

91 Methods

92 Study setting, laboratory procedures and DNA extraction

93 All isolates causing BSI between 15/09/2008-01/12/2018 (de-duplicated to one BSI/90-day
94 period within-patient) were processed by the clinical microbiology laboratory at the John
95 Radcliffe Hospital, Oxford, UK, using standard operating procedures, with sub-cultured stocks
96 stored at -80°C in 10% glycerol nutrient broth. The microbiology laboratory serves all healthcare
97 facilities in Oxfordshire, UK, with a catchment population of ~805,000. For sequencing, isolate
98 stocks were grown overnight on Columbia blood agar at 37°C and DNA was extracted using the
99 QuickGene DNA extraction kit (Autogen, MA, USA) as per the manufacturer's instructions with
100 the addition of a mechanical lysis step (FastPrep, MP Biomedicals, CA, USA; 6m/s for 40 secs).
101 Short-read sequencing was performed using the Illumina HiSeq 2500/3000/4000/MiSeq
102 instruments as previously described¹⁸. Additional, publicly available short-read sequencing data
103 was obtained to facilitate phylogenetic comparison with previous studies^{7,11,19,20} (BioProject
104 accessions: PRJNA480723, PRJEB4681, PRJEB35000, PRJEB12513).

105

106 Epidemiological linkage

107 Isolate data was linked to laboratory and electronic health record data via the Infections in
108 Oxfordshire Research Database (IORD). IORD has generic Research Ethics Committee, Health
109 Research Authority and Confidentiality Advisory Group approvals (19/SC/0403, 19/CAG/0144)
110 as a de-identified electronic research database. Data on suspected infectious focus and patient
111 provenance was acquired via linked local infection control records which had been submitted to
112 Public Health England as part of the mandatory surveillance programme; such data was
113 available for 400 *Klebsiella* spp. and 2773 *E. coli* isolates. Healthcare-associated (HA) BSI were

114 defined as occurring >48 hrs post-hospital admission or ≤30 days since hospital discharge;
115 other cases were defined as community-associated (CA)²¹.

116

117 **Genomic analysis**

118 All programs were run using default settings unless indicated. Raw reads were assembled using
119 Shovill (v1.0.4) with assemblies <4Mb and >7Mb excluded from further analysis, because these
120 were thought to represent possible assembly errors given the typical genome size for these
121 species (4-6.5Mb)²². De novo assemblies were annotated with Prokka (v1.14.6)²³; AMR genes,
122 virulence factors and plasmid replicons were identified using the Resfinder, VFDB, ISFinder and
123 PlasmidFinder databases with Abricate (v0.9.8) (--min-id 95 --min-cov 95)²⁴, and Kleborate for
124 *Klebsiella* spp.^{12,25-27}. Multi-locus sequence types (MLST) were determined *in silico* using the
125 MLST tool (v2.17.6) for *E. coli* and Kleborate for *Klebsiella* spp.^{28,29}. Phylogroups were
126 determined using the ClermonTyping tool³⁰. To fully utilise the resolution provided by whole
127 genome sequencing we additionally used fastBAPS³¹ to partition the population using the core
128 gene alignment produced by Panaroo³².

129

130 Mapping to MLST-specific reference genomes was performed using Snippy (v4.6.0)³³; genomes
131 were acquired from NCBI (appendix). Gubbins (v2.3.4) was used to build recombination-
132 corrected phylogenies³⁴. Time-scaled phylogenies were created using the BactDating library in
133 R (version 3.6.0) under a relaxed gamma model with a minimum of 100,000 Markov chain
134 Monte Carlo iterations³⁵. Significance of temporal signal was assessed with root-to-tip plots and
135 10,000 random permutations of tip dates. Effective sample size was assessed using the Coda
136 package³⁶. Evolutionary distinctiveness (ED) was calculated using the Picante package in R³⁷;
137 low ED scores indicate closely related genomes. All bioinformatics was performed using the
138 Oxford University Biomedical Research Computing Facility.

139

140 Contigs were classified as being of likely chromosomal or plasmid origin using MLplasmids with
141 a 0.7 probability cut-off³⁸. All plasmid/chromosomal contigs for a given isolate were binned into
142 separate multi-fasta files and the distances between these calculated using Dashing³⁹. The
143 pairwise distance matrix was filtered on a distance of 0.71 (the median plasmidome similarity of
144 bla_{CTX-M-15} carrying *K. pneumoniae* ST490 isolates) and then clustered into “plasmidome groups”
145 using the LinkComm package in R⁴⁰. The ecology of categorical groups was compared with a
146 permutational multivariate analysis of variance (permanova) test in the Vegan package in R⁴¹.

147

148 A pangenome wide association study (PGWAS) was conducted using PySeer⁴² with a linear
149 mixed model utilising relatedness distances inferred from a core genome phylogeny to correct
150 for population structure. A discriminant analysis of principal components (DAPC) was performed
151 using the Adegenet library in R⁴³. Gene co-occurrence was examined using graphs constructed
152 using the iGraph library in R⁴⁴. For each species, edge lists with all genes/plasmids/insertion
153 sequences with a Pearson correlation coefficient spp. >0.5 were created and communities
154 detected using single linkage.

155

156 **Statistical analysis**

157 To describe molecular epidemiological trends, STs were arbitrarily categorised as rare (≤ 10
158 isolates over the study period, or untypeable STs), intermediate (11-100 isolates), and for *E.*
159 *coli*, submajor (101-300 isolates) and major (≥ 300 isolates). For *Klebsiella* spp. we considered
160 the most prevalent ST (490) in isolation. Stacked negative binomial regression models with
161 clustered standard errors were used to compare rates of change over time (STATA v16⁴⁵)⁴⁶.
162 Incomplete years (2008 and 2018) are excluded from this part of the analysis. All other
163 statistical tests were performed in R v3.6 using the Stats package unless indicated⁴⁷. Charlson

164 score was calculated in the Comorbidity⁴⁸ R package using all ICD10 codes associated with
165 episodes in the year prior to specimen collection dates. To test for geographical/healthcare
166 setting structure in MLST groups we conducted a permutation test similar to that previously
167 described⁴⁹. In brief, the ratio of median distances between isolates from the same centre and
168 different centres was calculated. This observed ratio was compared to a permuted distribution
169 created by 1000 tip label randomisations. The number of permuted values at least as extreme
170 as the observed value divided by the number of permutations carried out was calculated to give
171 a one-sided test of statistical significance.

172 Results

173 **Escherichia coli bloodstream infections are stably dominated by major lineages across** 174 **community and healthcare-associated settings, but with diverse sporadic lineages** 175 **accounting for almost half of cases, and evidence of sub-lineage replacement in major** 176 **STs**

177 From September 2008-December 2018, 3461 *E. coli* isolates from 3196 patients were
178 sequenced. Major STs, namely ST73 (n=574), ST131 (457), ST95 (320) and ST69 (314)
179 comprised 48% of all isolates; sub-major STs, ST12 (144) and ST127 (113), 7% of all isolates;
180 intermediate STs (32 STs) 24% of all isolates; and rare STs (304 STs) 21% of all isolates
181 (Fig.1/Fig.S1). The incidence rate ratio per year (IRRY) of all these categories increased over
182 time (major IRRy=1.13 (95% CI:1.10-1.17), sub-major IRRy=1.16 (1.10-1.22), intermediate
183 IRRy=1.15 (1.12-1.18), rare IRRy=1.10 (1.06-1.14) (Fig.1A). There was no evidence that the
184 proportion of BSIs caused by major or sub-major STs changed over the study period (Fig.1B).
185 There was evidence that the incidence of BSIs caused by intermediate STs increased slightly
186 faster versus all other isolates and by rare isolates slightly slower ($p_{\text{heterogeneity}} < 0.05$). After sub-
8

187 stratifying the major STs 69, 73, 95 and 131 using BAPS, there was some evidence of
188 sublineage replacement (Fig.S2/Table.S1).

189
190 The majority of *E. coli* BSIs were CA-BSIs (2104/3461, 61%), with no evidence of variation in
191 the proportion caused by major STs between CA-BSI and HA-BSI (990/2104, 47%, vs
192 661/1357, 49%, $p=0.4$). Relatively few isolates came from patients resident in a nursing/care
193 home was relatively low (169 (7%) of 2301 with data available). The four major *E. coli* STs
194 accounted for a significantly greater proportion of these cases (97/169, 57% vs 1009/2132, 47%
195 $p=0.01$); however, there was no evidence of large-scale nursing/care home-associated BSI
196 outbreaks (Fig.S3). *E. coli* CA-BSIs were slightly older (median age 76 (IQR 63-85) vs 70 (55-
197 80), $p<0.001$) and less comorbid (median Charlson score 1 (IQR 0-2) vs 1 (IQR 1-2), $p=0.003$)
198 than HA-BSIs.

199
200 ***Klebsiella* spp. bloodstream infections are caused by a diverse representation of sub-**
201 **species with significant intra-species diversity, with the exception of *K. pneumoniae***
202 **ST490, causing a transient, clonal, local outbreak.**

203 Amongst the 886 successfully sequenced *Klebsiella* spp. isolates, a large number of sub-
204 species were observed, including *K. pneumoniae* (n=528 [60%]), *K. variicola* (n=112 [13%]), *K.*
205 *michiganensis* (n=90 [10%]), *K. oxytoca* (n=59 [7%]), *K. aerogenes* (n=38 [4%]), *K. grimontii*
206 (n=30 [3%]), *K. quasipneumoniae* (n=28 [3%]) and *K. africana* (n=1 [0.1%]). In stark contrast to
207 *E. coli*, 738/886 (83%) of *Klebsiella* spp. isolates belonged to rare STs (Fig.1/Fig.S4). The
208 multidrug resistant *K. pneumoniae* ST490 was the most prevalent *Klebsiella* ST and the only
209 one with >40 isolates; its incidence decreased over the study period (IRRY 0.78, 95%CI:0.68-
210 0.89). This ST is rarely seen in other studies, consistent with a transient but relatively large,
211 local outbreak. Notably 16/45 (36%) of isolates from this ST were community-onset.

212
213 The majority of *Klebsiella* spp. BSIs were HA-BSI (510/882 [missing data for 4 isolates], 59%).
214 As with *E. coli* few *K. pneumoniae* cases were in patients resident in a nursing/care home
215 (10/151, 7%) though this data was not available for most isolates. There was no difference in
216 the proportion of intermediate STs amongst HA-BSI vs CA-BSI (90/510 18% vs 58/372 16%,
217 $p=0.5$). *Klebsiella* spp. CA-BSI cases were older (median age 76 years (IQR 65-85) vs 66 years
218 (49-75) for HA-BSI; $p<0.001$) but also less comorbid than HA-BSI cases (median Charlson
219 score 1 (IQR 0-2) vs 2 (1-3); $p<0.001$). A large number (265/341, missing data for 31) of CA-
220 BSIs occurred in relatively healthy individuals (Charlson score ≤ 2 i.e. predicted 10 year survival
221 $\sim 90\%$). In contrast to previous reports suggesting *K. quasipneumoniae*/*K. variicola* may be less
222 virulent than *K. pneumoniae*⁷, in our study there was no evidence of differences in crude 30-day
223 mortality between the *Klebsiella* subspecies with ≥ 20 isolates (111/524 (21%) *K. pneumoniae*,
224 28/112 (25%) *K. variicola*, 23/89 (26%) *K. michiganensis*, 9/57 (16%) *K. oxytoca*, 10/36 (28%)
225 *K. aerogenes*, 7/28 (25%) *K. quasipneumoniae*, and 7/29 (24%) *K. grimontii*; exact $p=0.5$,
226 missing data for 9 isolates), nor the proportion of Ca-BSI (230/527 (44%), 48/112 (43%), 34/89
227 (38%), 27/58 (47%), 11/28 (39%), 9/27 (25%) and 12/30 (40%) respectively; exact $p=0.4$,
228 missing data for 4 isolates).

229
230 **No evidence of intra- or inter-regional clustering of major *Escherichia coli* lineages**
231 **causing BSI in either community- or healthcare-associated settings**

232 For the four largest *E. coli* STs, we compared the phylogeny of BSI isolates in this study to
233 those in previous studies over a similar timeframe (see Methods). There was no obvious

234 clustering of HA or CA isolates (Fig.2), supported by permutation testing, confirming that for the
235 major STs these were indeed randomly distributed across the phylogeny by geography and
236 healthcare/community setting (Table S2). Additionally, for all four of the major *E. coli* STs, there
237 was no difference in the ED scores between HA and CA cases, and therefore no evidence of
238 healthcare-associated adaptation and clonal expansion (Table S3). Our PGWAS did not identify
239 any significant hits separating CA and HA isolates (Fig.S5), including those that had been
240 previously identified in other studies⁴⁷.

241

242 **Urinary and hepatobiliary sources of *E. coli* and *Klebsiella* spp. BSI predominate, with**
243 **major *E. coli* STs over-represented in urinary-associated *E. coli* BSIs**

244 The urinary tract was the most common physician-identified source of infection in both species.
245 In *E. coli*, these were strongly associated with the presence of the pap group of proteins (Table
246 S4). The hepatobiliary was the second most common source followed by other gastrointestinal
247 infections for *E. coli* and a respiratory source for *Klebsiella* spp. (Fig.S6a). For *E. coli* BSIs,
248 there was some evidence that incidence of BSIs not attributed to the urinary tract increased
249 faster than those thought to be of urinary origin (IRRY 1.45, 95%CI 0.98-2.17 vs IRRy 1.36,
250 95%CI 0.88-2.08 respectively $p_{\text{heterogeneity}}=0.005$, Fig.S6b). The proportion of BSIs caused by the
251 five major *E. coli* STs was greater for BSIs with urinary compared to other sources (586/1071
252 [55%], vs 520/1229 [42%], $p<0.001$); this was not the case for *Klebsiella* spp. where the STs
253 causing BSI attributable to all sources were diverse.

254

255 **Multi-drug resistant *E. coli* BSIs are increasing, likely driven by the association of AMR**
256 **genes within lineages, and ceftriaxone resistance is driven by the expansion of resistant**
257 **ST131 and ST73 sub-clades**

258 For *E. coli* BSIs, considering total effects of resistance to individual antibiotics, regardless of
259 mechanism, there was evidence of increasing incidence of BSI carrying AMR genes conferring
260 resistance to amoxicillin IRRy=1.17 (95%CI 1.14-1.20), co-amoxiclav IRRy=1.15 (1.08-1.24),
261 ceftriaxone IRRy=1.26 (1.17-1.36), gentamicin IRRy=1.24 (1.15-1.33), trimethoprim IRRy=1.17
262 (1.14-1.20) and ciprofloxacin (IRRY=1.23 (1.17-1.29) Fig.3); there was only a single isolate
263 carrying a gene encoding for carbapenem resistance (ST10, bla_{OXA-48}) in 2014. For ceftriaxone
264 ($p_{\text{heterogeneity}}=0.003$), gentamicin ($p_{\text{heterogeneity}}=0.03$), amoxicillin ($p_{\text{heterogeneity}}=0.003$), ciprofloxacin
265 ($p_{\text{heterogeneity}}<0.001$), there was evidence that the incidence of isolates with AMR genes encoding
266 resistance to these antimicrobials was increasing faster than those without. Genes conferring
267 resistance to ceftriaxone in *E. coli* were over-represented in ST131 (180/346) and ST73
268 (56/346), with significantly lower ED scores for isolates carrying ceftriaxone-resistance genes in
269 these STs (accounting for 236/346 (68%) of these genes, Figs.S1), consistent with clonal
270 expansion of resistant sub-clades within these lineages. Findings were similar restricting to CA-
271 BSI. ST127 was the only major/sub-major ST in which no genes conferring resistance to
272 ceftriaxone were detected.

273
274 **Drug-susceptible *Klebsiella* spp. BSIs are increasing faster than drug-resistant strains,**
275 **with ceftriaxone-resistance largely healthcare-associated**

276 For *Klebsiella* spp., there was only evidence of increasing incidence of isolates carrying
277 genes/mutations conferring resistance to co-amoxiclav IRRy=1.07 (95%CI:1.02-1.12) and
278 ciprofloxacin IRRy=1.11 (1.06-1.16) (Fig.3). In contrast, the incidence of genes conferring
279 resistance to ceftriaxone IRRy 0.99 (0.92-1.09) and trimethoprim IRRy=1.03 (0.93-1.15) were
12

280 stable while gentamicin IRRy=0.86 (0.76-0.97) decreased. For ceftriaxone, co-amoxiclav and
281 gentamicin, the incidence trends of isolates not carrying genes conferring resistance to these
282 antibiotics increased faster than those carrying resistance ($p_{\text{heterogeneity}} \leq 0.01$).

283
284 A major contributing factor to this finding was the overall decline of the MDR ST490 lineage and
285 the emergence of a more susceptible ST490 sub-lineage between 2005-2010 which had lost the
286 *aac(3)-IIa*, *aac(6')-Ib-cr*, *bla_{OXA-1}* and *tet(a)* genes (Fig.S7). Genes encoding for resistance to
287 ceftriaxone were significantly more common in intermediate STs (STs occurring >10 times in the
288 dataset) (54/148, 36% vs 73/738, 10%, $p < 0.001$) and in HA-BSIs (85/510, 17% vs 42/372, 11%
289 $p=0.03$). Similarly MDR isolates (resistant to ≥ 3 antibiotic classes) were more common in
290 intermediate vs rare STs (89/366, 24% vs 80/520, 15%, $p=0.001$) and HA-BSI vs CA-BSI
291 (109/510, 21% vs 59/372, 16%, $p=0.049$).

292
293 **Virulence factors are strongly structured by ST amongst E. coli but not Klebisella spp.**
294 **BSI isolates with no evidence in Klebsiella of a difference in virulence gene carriage**
295 **between community vs hospital acquired BSIs.**

296 The distribution of virulence factors strongly reflected the underlying clonal population structure
297 in *E. coli* isolates, with segregation by ST, confirmed with a discriminant analysis of principal
298 components (DAPC) (Fig.4). Most notably phylogroup B2 isolates were separated from the rest
299 of the population by the presence of *chuA*, *chuX* (involved in haem utilisation) and *espL1/espX4*
300 (elements of the type 3 secretion system). Given this and the broad equilibrium in the *E. coli* BSI
301 population structure demonstrated above, it is unlikely that increased carriage of certain
302 virulence factors explains the increasing incidence of these infections. In keeping with this, there
303 was a near-perfect correlation between frequency of carriage of any given virulence gene
304 across all isolates in 2009 vs 2018 (Pearson correlation=0.99; $p < 0.001$).

305
306 No such structuring by virulence factors was observed for *Klebsiella* spp.; instead, virulence
307 factors were widely distributed amongst 79 known STs and 14 isolates with no assigned ST
308 (Fig.S8). Notably, of the 372 CA-BSIs, only 65 (17%) carried known virulence factors. There
309 was no difference in the proportion of CA-BSI and HA-BSI isolates carrying ≥ 1 virulence factor
310 (65/372 [17%] vs 92/510 [18%] respectively; $p=0.9$) nor carrying colibactin and/or aerobactin
311 (15/372 [4%] vs 25/510 [5%]; $p=0.7$) (Fig.5).

312
313 **Plasmid replicon and plasmidome analyses reflect diversity consistent with a highly**
314 **mobile accessory genome within species**

315 Given the known importance of plasmids in the carriage and transmission of AMR genes, we
316 sought to understand the role they might play in the relative success of major STs. Analysis of
317 plasmid replicon profiles showed these were largely genus-restricted, with some overlap
318 (Fig.S9). However, plasmid replicons were not structured by host lineage in *E. coli* or *Klebsiella*
319 spp., with the exception of *K. pneumoniae* ST490 which was discriminated by the presence of
320 the IncFIA(HI1) plasmid type (also identified in 23 *E. coli* isolates; Fig.S10).

321
322 Using our approach to predict plasmidome population structure based on k-mer similarity of
323 contigs binned as plasmid, we observed a striking degree of plasmidome diversity within highly
324 genomically-related isolates/STs for all species (Fig.S9); for 656 isolates with a core genome
325 similarity >0.99 to another isolate in the dataset, the median plasmidome similarity was only
326 0.51 (IQR 0.28-0.86). Isolates with a near-identical plasmidome (>0.99 similarity, $n=115$
327 isolates) did however have highly similar chromosomes (median plasmidome similarity: 0.97

328 [IQR: 0.94-0.99]). Compared with other STs, major *E. coli* STs had larger plasmidomes (median
329 size 106,766 vs 97,432, $p < 0.001$) which belonged to more “plasmidome groups” (see methods)
330 (median 3 vs 2, $p < 0.001$). For both *E. coli* and *Klebsiella* spp., there was no evidence of
331 different plasmid populations between those associated with HA- vs CA-BSI ($p = 0.4$).
332
333 Finally we analysed the co-occurrence of AMR genes, plasmid replicon types and insertion
334 sequences (ISs) within *E. coli* and *Klebsiella* spp. The networks formed were notable for the
335 widespread co-occurrence of AMR genes and insertion sequences but not usually specific
336 plasmid types (Figs.S11/12). This suggests that most common AMR genes are found in a
337 diverse range of genetic contexts (e.g. multiple plasmid types, chromosomally integrated), and
338 that horizontal gene transfer of important AMR genes in these isolates is largely facilitated by
339 smaller, non-plasmid MGEs such as transposons/ISs, which are difficult to reliably evaluate with
340 short-read data⁵⁰.

341 Discussion

342 In this unbiased longitudinal sequencing study of all (90 day-deduplicated) *E. coli* and *Klebsiella*
343 spp. BSIs in Oxfordshire over a decade (2008-2018), we highlight the similarities and
344 differences in the molecular epidemiology of these species. Overall, the increasing incidence of
345 *E. coli* and *Klebsiella* spp. BSIs in Oxfordshire is not explained by the expansion of a single ST,
346 and much of the burden (40%) of *Klebsiella* spp. disease is caused by non-*K. pneumoniae*
347 species. Although six lineages stably account for nearly half of all *E. coli* BSIs and a clonal
348 outbreak was observed in *K. pneumoniae*, many BSIs in our setting are caused by diverse
349 strains with diverse accessory genomes.

350

351 We found no genomic evidence supporting the stratification of *E. coli* isolates into HA-BSI vs
352 CA-BSI. Strains causing healthcare-associated BSIs may not be specifically healthcare-
353 acquired but may still be healthcare-provoked (for example by the presence of indwelling
354 devices or relative immunosuppression). Interventions targeting only HA infections (such as that
355 of the UK government to reduce these by 50% by 2021¹⁴) might have limited efficacy without
356 considering the wider ecology of these species. Whilst the incidence of *E. coli* BSIs attributed to
357 the urinary tract increased over the study period, there was evidence that the rate of this
358 increase was slower compared to other sources. This possibly reflects some success of
359 infection prevention interventions such as catheter care, but also highlights the multi-faceted
360 approach required to reduce the overall incidence rate. More work is required to understand CA-
361 *Klebsiella* spp. BSIs because known markers of virulence found in CA disease (e.g. aerobactin,
362 yersiniabactin, colibactin) in other studies⁴⁸ were not seen in most of our cases. Importantly
363 therefore, existing markers may therefore be insensitive to detect the emergence of clinically
364 hypervirulent, CA, multidrug-resistant strains.

365
366 For *E. coli*, our analysis is consistent with previous studies demonstrating a broadly stable
367 population structure at the MLST level with the predominance of STs 69, 73, 95 and 131. Within
368 this however, we demonstrated some replacement in sub-clades of major STs over the study
369 period. This may represent an adaptive strategy within major STs that allows them to continue
370 to maintain their niche and evolve to survive changing environmental and immunological
371 selection pressures. Differences in the presence of haem utilisation genes was a notable factor
372 which discriminated isolates in major STs from the rest of the population, suggesting that iron
373 metabolism might be an important selective pressure in determining invasive potential, a finding
374 supported by a recent GWAS of a mouse model⁵¹. In contrast, for *Klebsiella* spp. we

375 demonstrated that most infections are caused by sporadic STs which individually only
376 accounted for only a small proportion of the overall population causing disease.

377

378 In *E. coli*, the incidence of isolates carrying genes conferring resistance against commonly used
379 antimicrobial classes increased faster than those without; in general the opposite was true for
380 *Klebsiella* spp.. The most commonly isolated MDR-ST in *Klebsiella* spp. (ST490) significantly
381 decreased in incidence over the study period, lending credence to the idea that the relentless
382 expansion of MDR clades is not inevitable, although it is unclear what interventions may have
383 helped in causing its decline, particularly given that approximately a third of cases were
384 community onset. The increasing incidence of gentamicin/ceftriaxone resistance in
385 predominantly CA-*E. coli* BSIs seems paradoxical given that these antibiotics are rarely used in
386 this setting. One explanation is that they are co-located on MGEs with other AMR genes
387 encoding resistance to antibiotics more commonly used in the community (e.g. amoxicillin,
388 trimethoprim). An alternative hypothesis might be increased exposure to third-generation
389 cephalosporins due to the rise of ambulatory/“hospital-at-home” medical pathways where the
390 once-a-day dosing of ceftriaxone makes it a relatively widely prescribed antibiotic in these
391 settings. Regardless, overall, *Klebsiella* strains causing BSI appear to be exposed to declining
392 antibiotic selection pressures compared to *E. coli*, suggesting that they are maintained and
393 selected for in distinct ecological niches.

394

395 In *E. coli* BSI isolates, there was no genetic signal of adaptation to either healthcare or
396 community environments, suggesting that these are not relevant niches for selection, contrary to
397 a previous study⁴⁹. However, this study included small numbers of isolates (n=162), and may
398 have been unable to fully account for population structure. Furthermore our analysis suggests
399 that, contrary to true nosocomial pathogens⁵², the ecology of the *E. coli* plasmidome is similar
17

400 for HA/CA infections. Our findings support the hypothesis that the diversity observed in any
401 given epidemiological strata (eg. age groups/infection focus/healthcare setting) is sampled from
402 the same common ecological pool of isolates with invasive potential rather than representing
403 specialised adaptation to any given setting. There are of course exceptions to this, such as
404 localised outbreaks within care facilities and hospital environments.⁵³ However, our data
405 suggests that at least in Oxfordshire, these outbreaks are do not contribute significantly to the
406 epidemiology of *E. coli* BSIs. Importantly, iatrogenic, non-pathogen-associated factors
407 promoting invasive infection (e.g. urinary catheterisation) should be minimised to reduce the
408 incidence of *E. coli* BSIs.

409
410 For *Klebsiella* spp., the traditional view that HA infections are opportunistic and caused by MDR
411 isolates and CA infections are caused by isolates carrying one or more specific virulence genes
412 (eg yersiniabactin, colibactin, aerobactin, salmochelin) appears to be an over-simplification in
413 our setting^{12,13}. The majority of CA *Klebsiella* spp. infections contained none of the known
414 genetic markers of hypervirulence, and these genes are unlikely to be a reliable method of
415 surveillance for emerging strains with a propensity to cause invasive CA disease, at least in our
416 setting. Whilst ESBL and MDR isolates were significantly more common in HA isolates, about a
417 third were community onset (even using our fairly conservative definition) which significantly
418 challenges the prevailing dogma that these are largely opportunistic HA strains.

419
420 Our study was limited by the inability to reconstruct closed genomes/plasmids using short-read
421 sequencing data and our plasmidome analysis should therefore be interpreted with caution. The
422 widely used PlasmidFinder database allows some inferences to be made using short read data,
423 however there are many untypeable plasmids that are not reflected in this database. Similarly
424 the MLplasmids classification algorithm used is trained on a limited reference set which may

425 lead to some erroneous contig classifications. Additionally, analysis of the entire plasmidome
426 may be too crude to identify the nuances of similarities/differences, particularly for small, shared
427 plasmids as these represent only a small proportion of the overall plasmidome. Whilst we used
428 a relatively conservative definition of healthcare-associated (i.e. within 30 days of hospital
429 admission), this may have failed to capture the longer-term impacts of earlier hospital
430 admissions.

431
432 In summary, the contrasting epidemiology of *E. coli* and *Klebsiella spp.* BSIs suggests different
433 reservoirs, selection pressures and modes of acquisition for these genera. Separate strategies
434 to reduce the incidence of these infections are likely to be required, and consideration of the
435 community as a reservoir is important. The lack of reproducibility of several findings from
436 previous studies and the poor sensitivity of current molecular markers for hyper-virulent
437 *Klebsiella* surveillance highlights the critical importance of unselected sampling frames when
438 making epidemiological inferences.

439 Authors Notes

440 **Author Contributions:**

441 ASW, NS, DC and TEAP planned and managed isolate collection and acquired funding for the
442 study. SL and KDV performed data linkage, cleaning and analysis. SL wrote the first draft of the
443 manuscript. KC, LB, SG, JK, TD and AV performed the laboratory work. KJ and MA facilitated
444 access to isolates and resources. ASW, SH, NS, DC and TEAP provided supervision. All
445 authors reviewed and approved the final draft for publication.

446

447 **Declaration of interests:**

448 The authors declare that there are no relevant interests.

449

450 **Data availability:**

451 All sequencing data has been deposited in the NCBI under project accession number

452 PRJNA604975.

453

454 **Funding:**

455 The research was supported by the National Institute for Health Research (NIHR)

456 Health Protection Research Unit in Healthcare Associated Infections and Antimicrobial

457 Resistance (NIHR200915) at the University of Oxford in partnership with Public

458 Health England (PHE) and by Oxford NIHR Biomedical Research Centre. T Peto and

459 AS Walker are NIHR Senior Investigators. The report presents independent research

460 funded by NIHR. The views expressed in this publication are those of the authors and

461 not necessarily those of the NHS, NIHR, the Department of Health or Public Health

462 England. The computational aspects of this research were funded from the NIHR Oxford

463 BRC with additional support from the Wellcome Trust Core Award Grant Number

464 203141/Z/16/Z. SL is supported by a Medical Research Council Clinical Research Training

465 Fellowship. KC is Medical Research Foundation-funded.

466

467 **Acknowledgements:**

468 We express our thanks to colleagues in the Oxford University Hospitals NHS Foundation Trust

469 clinical microbiology laboratory who performed routine identification of the isolates in this study

470 and catalogued then for storage. We are also indebted to Dai Griffiths who managed the isolate
471 collected for many years. This work uses data provided by patients and collected by the UK's
472 National Health Service as part of their care and support. We thank all the people of Oxfordshire
473 who contribute to the Infections in Oxfordshire Research Database.
474 Research Database Team: L Butcher, H Boseley, C Crichton, DW Crook, D Eyre, O Freeman, J
475 Gearing (community), R Harrington, K Jeffery, M Landray, A Pal, TEA Peto, TP Quan, J
476 Robinson (community), J Sellors, B Shine, AS Walker, D Waller. Patient and Public Panel: G
477 Blower, C Mancey, P McLoughlin, B Nichols.

478 References

- 479
- 480
- 481 1. Vihta, K.-D. *et al.* Trends over time in Escherichia coli bloodstream infections, urinary
482 tract infections, and antibiotic susceptibilities in Oxfordshire, UK, 1998–2016: a study of
483 electronic health records. *Lancet Infect. Dis.* **18**, 1138–1149 (2018).
 - 484 2. *English Surveillance Programme for Antimicrobial Utilisation and Resistance (ESPAUR)*.
485 [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/fi](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/843129/English_Surveillance_Programme_for_Antimicrobial_Utilisation_and_Resistance_2019.pdf)
486 [le/843129/English_Surveillance_Programme_for_Antimicrobial_Utilisation_and_Resistance_20](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/843129/English_Surveillance_Programme_for_Antimicrobial_Utilisation_and_Resistance_2019.pdf)
487 [19.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/843129/English_Surveillance_Programme_for_Antimicrobial_Utilisation_and_Resistance_2019.pdf) (2019).
 - 488 3. Public Health England. Escherichia coli (E. coli): guidance, data and analysis. *GOV.UK*
489 <https://www.gov.uk/government/collections/escherichia-coli-e-coli-guidance-data-and-analysis>
490 (2010).
 - 491 4. Stoesser, N. *et al.* Evolutionary History of the Global Emergence of the Escherichia coli
492 Epidemic Clone ST131. *MBio* **7**, e02162 (2016).

- 493 5. David, S. *et al.* Epidemic of carbapenem-resistant *Klebsiella pneumoniae* in Europe is
494 driven by nosocomial spread. *Nat Microbiol* **4**, 1919–1929 (2019).
- 495 6. Wyres, K. L. *et al.* Distinct evolutionary dynamics of horizontal gene transfer in drug
496 resistant and virulent clones of *Klebsiella pneumoniae*. *PLoS Genet.* **15**, e1008114 (2019).
- 497 7. Goswami, C. *et al.* Genetic analysis of invasive *Escherichia coli* in Scotland reveals
498 determinants of healthcare-associated versus community-acquired infections. *Microb Genom* **4**,
499 (2018).
- 500 8. Blandy, O. *et al.* Factors that impact on the burden of *Escherichia coli* bacteraemia:
501 multivariable regression analysis of 2011-2015 data from West London. *J. Hosp. Infect.* **101**,
502 120–128 (2019).
- 503 9. Kizny Gordon, A. *et al.* Genomic dynamics of species and mobile genetic elements in a
504 prolonged blaIMP-4-associated carbapenemase outbreak in an Australian hospital. *J.*
505 *Antimicrob. Chemother.* **75**, 873–882 (2020).
- 506 10. Day, M. J. *et al.* Population structure of *Escherichia coli* causing bacteraemia in the UK
507 and Ireland between 2001 and 2010. *J. Antimicrob. Chemother.* **71**, 2139–2142 (2016).
- 508 11. Kallonen, T. *et al.* Systematic longitudinal survey of invasive *Escherichia coli* in England
509 demonstrates a stable population structure only transiently disturbed by the emergence of
510 ST131. *Genome Res.* (2017) doi:10.1101/gr.216606.116.
- 511 12. Lam, M. M. C. *et al.* Tracking key virulence loci encoding aerobactin and salmochelin
512 siderophore synthesis in *Klebsiella pneumoniae*. *Genome Med.* **10**, 77 (2018).
- 513 13. Holt, K. E. *et al.* Genomic analysis of diversity, population structure, virulence, and
514 antimicrobial resistance in *Klebsiella pneumoniae*, an urgent threat to public health. *Proc. Natl.*
515 *Acad. Sci. U. S. A.* **112**, E3574-81 (2015).
- 516 14. Wyres, K. L. *et al.* Genomic surveillance for hypervirulence and multi-drug resistance in
517 invasive *Klebsiella pneumoniae* from South and Southeast Asia. *Genome Med.* **12**, 11 (2020).

- 518 15. Long, S. W. *et al.* Population Genomic Analysis of 1,777 Extended-Spectrum Beta-
519 Lactamase-Producing *Klebsiella pneumoniae* Isolates, Houston, Texas: Unexpected
520 Abundance of Clonal Group 307. *MBio* **8**, (2017).
- 521 16. SURVEILLANCE REPORT. Surveillance of antimicrobial resistance in Europe 2018.
- 522 17. Ellington, M. J. *et al.* Contrasting patterns of longitudinal population dynamics and
523 antimicrobial resistance mechanisms in two priority bacterial pathogens over 7 years in a single
524 center. *Genome Biol.* **20**, 184 (2019).
- 525 18. De Maio, N. *et al.* Comparison of long-read sequencing technologies in the hybrid
526 assembly of complex bacterial genomes. *Microb Genom* **5**, (2019).
- 527 19. Hastak, P. *et al.* Genomic profiling of *Escherichia coli* isolates from bacteraemia patients:
528 a 3-year cohort study of isolates collected at a Sydney teaching hospital. *Microb Genom* **6**,
529 (2020).
- 530 20. van Hout, D. *et al.* Extended-spectrum beta-lactamase (ESBL)-producing and non-
531 ESBL-producing *Escherichia coli* isolates causing bacteremia in the Netherlands (2014 - 2016)
532 differ in clonal distribution, antimicrobial resistance gene and virulence gene content. *PLoS One*
533 **15**, e0227604 (2020).
- 534 21. *Guidance on the definition of healthcare associated Gram-negative bloodstream*
535 *infections*. [https://www.england.nhs.uk/wp-](https://www.england.nhs.uk/wp-content/uploads/2020/08/HCA_BSI_definitions_guidance.pdf)
536 [content/uploads/2020/08/HCA_BSI_definitions_guidance.pdf](https://www.england.nhs.uk/wp-content/uploads/2020/08/HCA_BSI_definitions_guidance.pdf).
- 537 22. Seemann, T. Shovill. <https://github.com/tseemann/shovill>.
- 538 23. Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068–
539 2069 (2014).
- 540 24. Seemann, T. *abricate*. (Github).

- 541 25. Lam, M. M. C. *et al.* Genetic diversity, mobilisation and spread of the yersiniabactin-
542 encoding mobile element ICEKp in *Klebsiella pneumoniae* populations. *Microb Genom* **4**,
543 (2018).
- 544 26. Wyres, K. L. *et al.* Identification of *Klebsiella* capsule synthesis loci from whole genome
545 data. *Microb Genom* **2**, e000102 (2016).
- 546 27. Wick, R. R., Heinz, E., Holt, K. E. & Wyres, K. L. Kaptive Web: User-Friendly Capsule
547 and Lipopolysaccharide Serotype Prediction for *Klebsiella* Genomes. *J. Clin. Microbiol.* **56**,
548 (2018).
- 549 28. Seemann, T. *mlst*. (Github).
- 550 29. Holt, K. Kleborate. <https://github.com/katholt/Kleborate>.
- 551 30. Beghain, J., Bridier-Nahmias, A., Le Nagard, H., Denamur, E. & Clermont, O.
552 ClermonTyping: an easy-to-use and accurate in silico method for *Escherichia* genus strain
553 phylotyping. *Microb Genom* **4**, (2018).
- 554 31. Tonkin-Hill, G., Lees, J. A., Bentley, S. D., Frost, S. D. W. & Corander, J. Fast
555 hierarchical Bayesian analysis of population structure. *Nucleic Acids Res.* **47**, 5539–5549
556 (2019).
- 557 32. Tonkin-Hill, G. *et al.* Producing Polished Prokaryotic Pangenomes with the Panaroo
558 Pipeline. *bioRxiv* 2020.01.28.922989 (2020) doi:10.1101/2020.01.28.922989.
- 559 33. Seemann, T. *snippy*. (Github, 2015).
- 560 34. Croucher, N. J. *et al.* Rapid phylogenetic analysis of large samples of recombinant
561 bacterial whole genome sequences using Gubbins. *Nucleic Acids Res.* **43**, e15 (2015).
- 562 35. Didelot, X., Croucher, N. J., Bentley, S. D., Harris, S. R. & Wilson, D. J. Bayesian
563 inference of ancestral dates on bacterial phylogenetic trees. *Nucleic Acids Res.* **46**, e134
564 (2018).

- 565 36. Plummer, M., Best, N., Cowles, K. & Vines, K. CODA: Convergence Diagnosis and
566 Output Analysis for MCMC. *R News* vol. 6 7–11 (2006).
- 567 37. Kembel, S. W. *et al.* Picante: R tools for integrating phylogenies and ecology.
568 *Bioinformatics* **26**, 1463–1464 (2010).
- 569 38. Arredondo-Alonso, S. *et al.* mlplasmids: a user-friendly tool to predict plasmid- and
570 chromosome-derived sequences for single species. *Microb Genom* **4**, (2018).
- 571 39. Baker, D. N. & Langmead, B. Dashing: fast and accurate genomic distances with
572 HyperLogLog. *Genome Biol.* **20**, 265 (2019).
- 573 40. Kalinka, A. T. & Tomancak, P. linkcomm: an R package for the generation, visualization,
574 and analysis of link communities in networks of arbitrary size and type. *Bioinformatics* **27**, 2011–
575 2012 (2011).
- 576 41. Oksanen, J. *et al.* vegan: Community Ecology Package. (2019).
- 577 42. Lees, J. A., Galardini, M., Bentley, S. D., Weiser, J. N. & Corander, J. pyseer: a
578 comprehensive tool for microbial pangenome-wide association studies. *Bioinformatics* **34**,
579 4310–4312 (2018).
- 580 43. Jombart, T. & Ahmed, I. adegenet 1.3-1: new tools for the analysis of genome-wide SNP
581 data. *Bioinformatics* **27**, 3070–3071 (2011).
- 582 44. Csardi, G. & Nepusz, T. The igraph software package for complex network research.
583 *InterJournal* vol. Complex Systems 1695 (2006).
- 584 45. StataCorp, L. P. *Stata statistical software: release 16. College Station, TX: StataCorp*
585 *LP; 2019.*
- 586 46. Lunn, M. & McNeil, D. Applying Cox regression to competing risks. *Biometrics* **51**, 524–
587 532 (1995).
- 588 47. R Core Team. R: A Language and Environment for Statistical Computing. (2019).

- 589 48. Gasparini, A. comorbidity: An R package for computing comorbidity scores. *Journal of*
590 *Open Source Software* **3**, 648 (2018).
- 591 49. Eyre, D. W. *et al.* Two Distinct Patterns of Clostridium difficile Diversity Across Europe
592 Indicating Contrasting Routes of Spread. *Clin. Infect. Dis.* **67**, 1035–1044 (2018).
- 593 50. Wick, R. R., Judd, L. M., Gorrie, C. L. & Holt, K. E. Completing bacterial genome
594 assemblies with multiplex MinION sequencing. *Microb Genom* **3**, e000132 (2017).
- 595 51. Galardini, M. *et al.* Major role of iron uptake systems in the intrinsic extra-intestinal
596 virulence of the genus Escherichia revealed by a genome-wide association study. *PLoS Genet.*
597 **16**, e1009065 (2020).
- 598 52. Arredondo-Alonso, S. *et al.* Plasmids Shaped the Recent Emergence of the Major
599 Nosocomial Pathogen Enterococcus faecium. *MBio* **11**, (2020).
- 600 53. Brodrick, H. J. *et al.* Longitudinal genomic surveillance of multidrug-resistant Escherichia
601 coli carriage in a long-term care facility in the United Kingdom. *Genome Med.* **9**, 70 (2017).

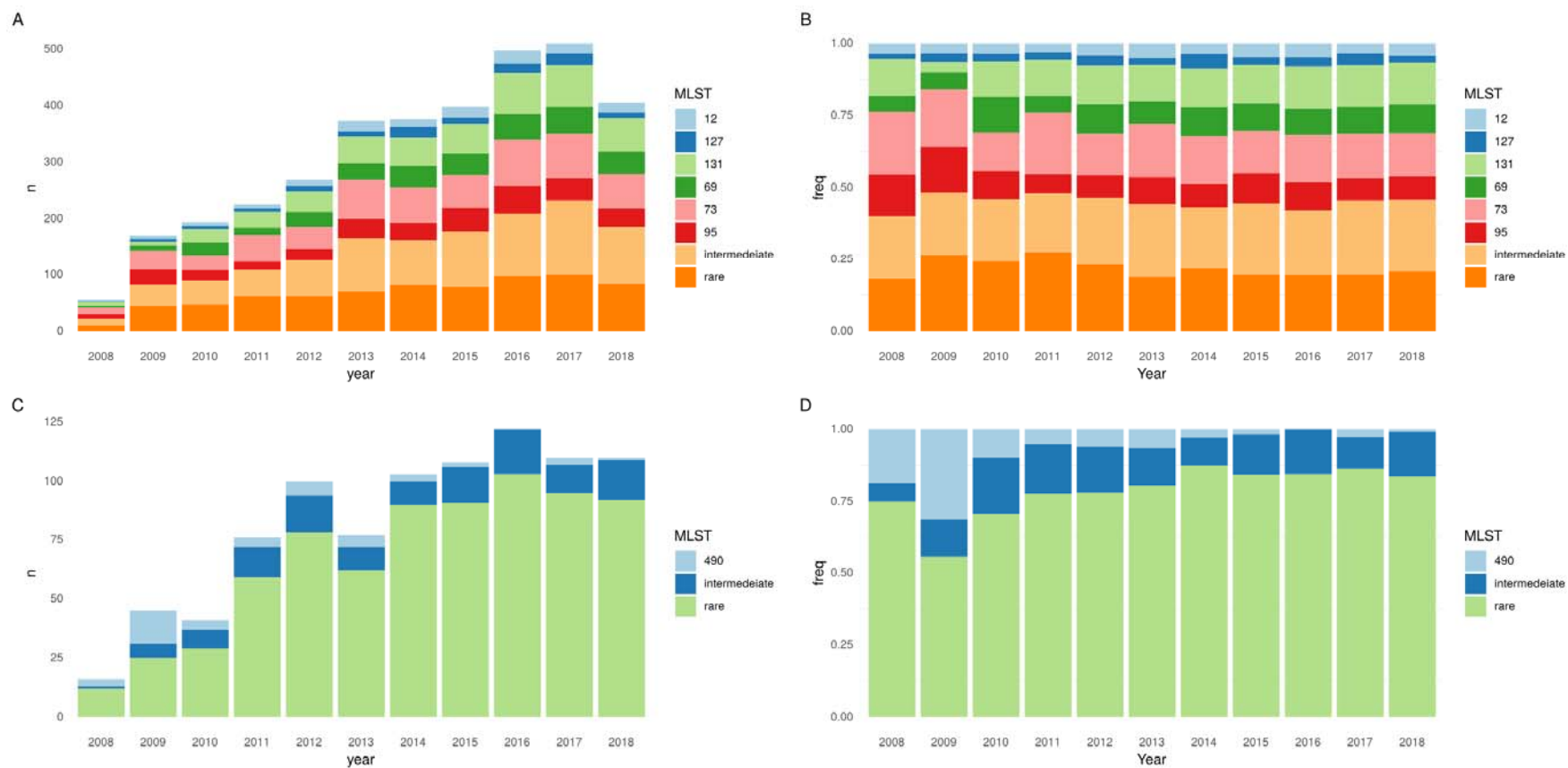


Figure 1 - Population dynamics of *E. coli* (A, B) and *Klebsiella* spp. (C, D) STs over time. A) Absolute number and B) proportions of *E. coli* BSIs caused by four major STs (STs 131/95/73/69), sub-major STs (STs 127/12), intermediate STs and rare STs. C) Absolute number and D) proportions of *Klebsiella* spp. BSIs caused by ST490, intermediate, and rare STs. N.B years 2008 and 2018 are incomplete (see methods).

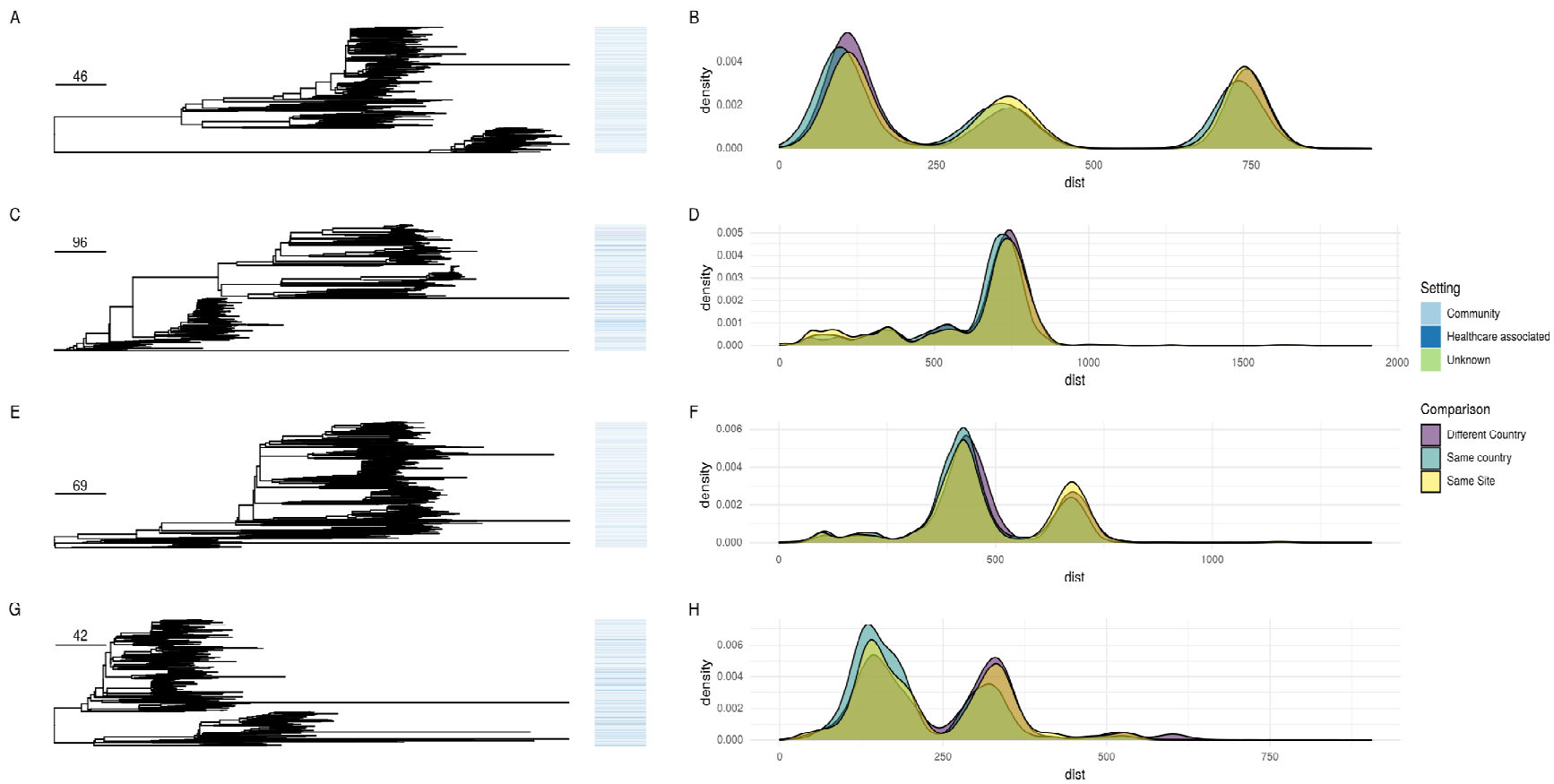


Figure 2: Left panel: Core genome phylogenies (corrected for recombination) for ST131/95/73/69 (top to bottom). Origin of isolates is denoted by the colour in the bar to the right of the tree. Scale bar shows SNPs. Right panel shows the distribution of cophenetic distances for the trees shown for isolates from the same site, same country, different countries.

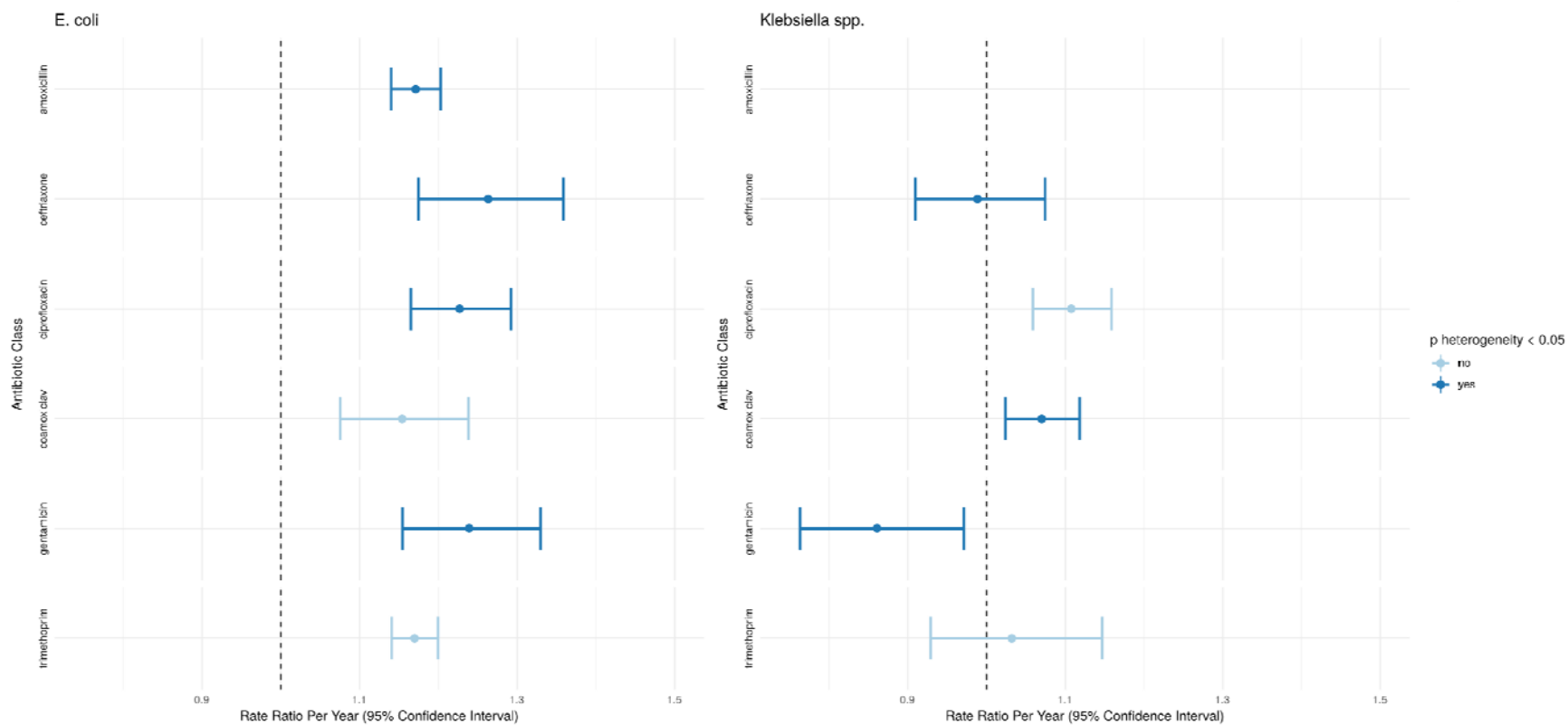


Figure 3: Changes in the incidence of clinically relevant groups of antimicrobial resistance genes considering total effects of resistance to individual antibiotics, regardless of mechanism, over the study period. Gene groupings are as defined by ResFinder. Amoxicillin is excluded for *Klebsiella spp.* which are generally considered intrinsically resistant to this drug. Bars show univariate point estimates and 95% confidence intervals for the incidence rate ratio (per year) of presence of genes belonging to the classes shown. The point estimate and error bar is shown in a darker blue if a Wald test between the incidence rate ratios for the presence and absence of a given set of genes was significant at a 0.05 threshold (i.e. there was evidence that the rate of increase of isolates with a group of genes was different to that for isolates without.)

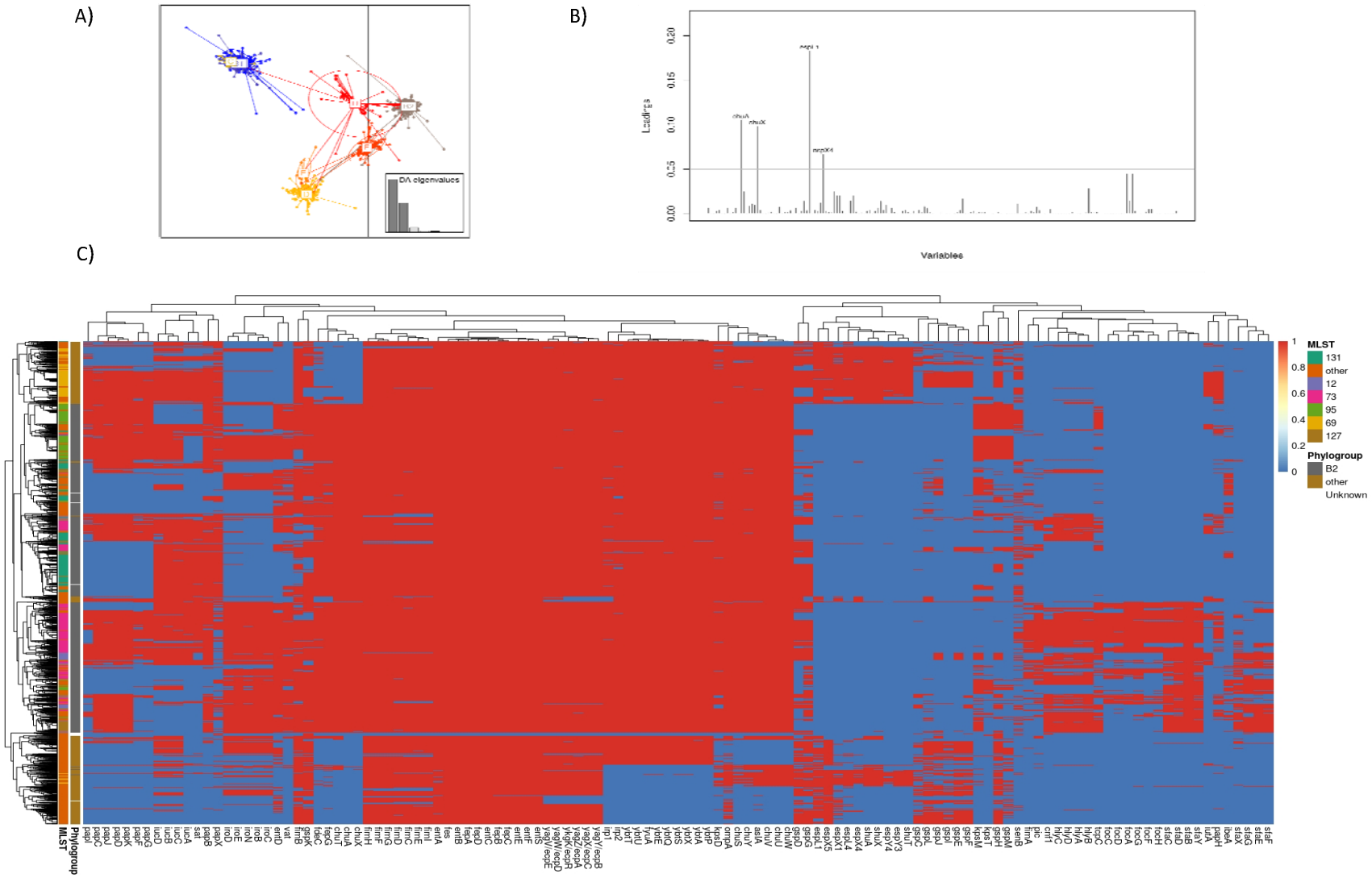


Figure 4: Virulence elements are structured by ST/Phylogroup in E. coli. A) Discriminant analysis of principal components plot showing clear separation of phylogroups by their virulence factor content. B) Loading plot showing genes contributing most to the discrimination of phylogroup B2 from other phylogroups. C) Heatmap of presence absence of virulence genes (x-axis) by major ST and phylogroups (y-axis).

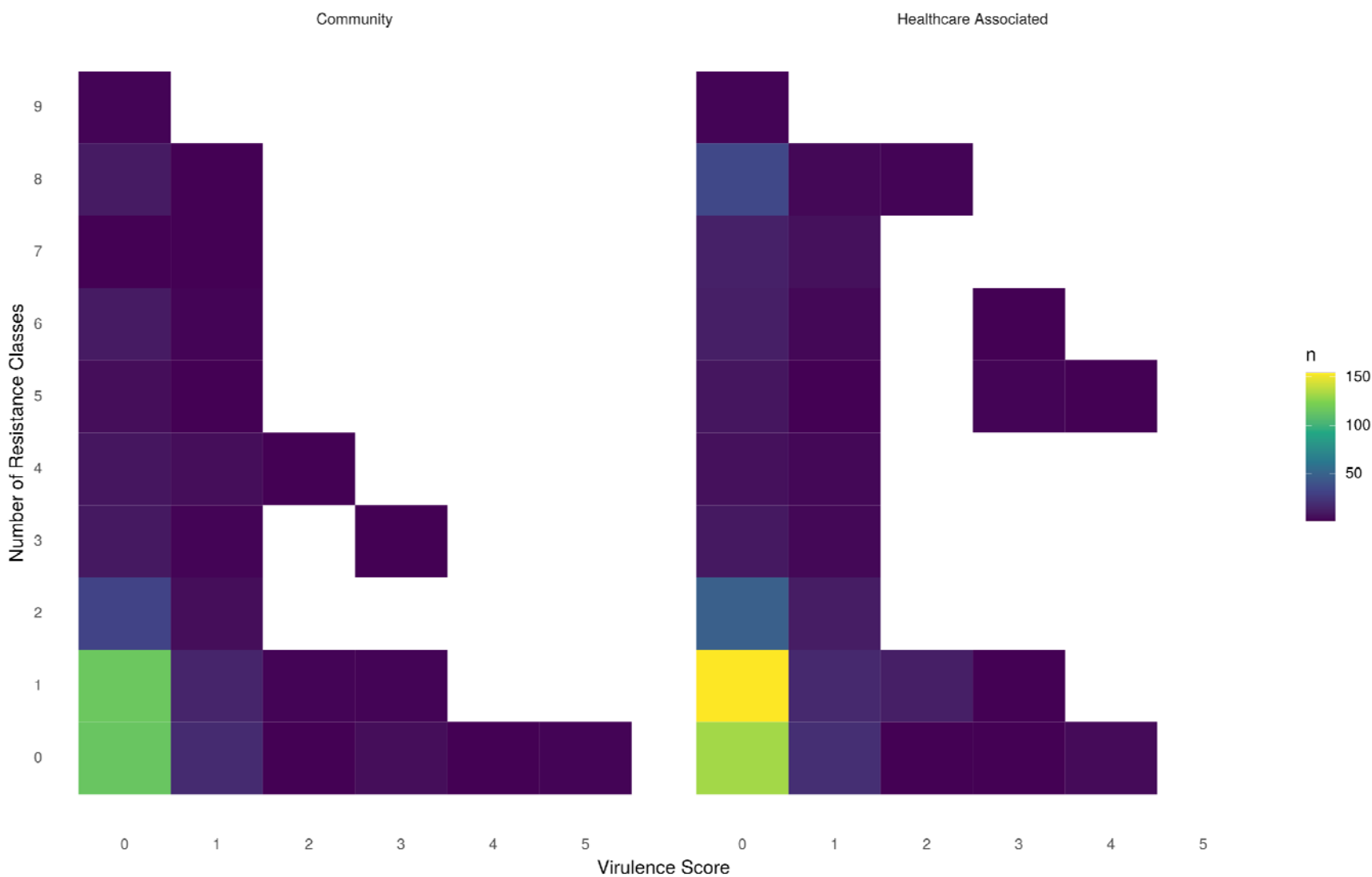


Figure 5: Distribution of virulence factors and antimicrobial resistance classes across community acquired (left) and healthcare associated (right) *Klebsiella* spp. isolates. Virulence factor scores were assigned by Kleborate as follows: Virulence Score 0 = No acquired loci, 1= yersiniabactin, 2= yersiniabactin and colibactin (or colibactin only = aerobactin (without yersiniabactin or colibactin), 4 = aerobactin with yersiniabactin (without colibactin) and 5 = yersiniabactin, colibactin and aerobactin.