Multi-omic surveillance of *Escherichia coli* and *Klebsiella* spp. in hospital sink drains and patients

Constantinides B,[a,b,c,†,*] **Chau KK**,[a,b,c,†] **Quan TP**,[a,b,c] **Rodger G**,[a,b,c] **Andersson M**,[d] **Jeffery K**,[d] **Lipworth S**,[a] **Gweon HS**,[e] **Peniket A**,[f] **Pike G**,[g] **Millo J**,[h] **Byukusenge M**,[i] **Holdaway M**,[h] **Gibbons C**,[g] **Mathers A J**,[i,j] **Crook DW**,[a,b,c,d] **Peto TEA**,[a,b,c,d] **Walker AS**,[a,b,c] **Stoesser N**[a,b,c,d]

Nuffield Department of Medicine, University of Oxford, Oxford, UK [a] National Institute for Health Research (NIHR) Health Protection Research Unit on Healthcare Associated Infections and Antimicrobial Resistance, John Radcliffe Hospital, Oxford, UK [b] NIHR Biomedical Research Centre, Oxford, UK [c] Department of Infectious diseases/Microbiology, Oxford University Hospitals NHS Foundation Trust, Oxford, UK [d] Harborne Building, School of Biological Sciences, University of Reading, Reading, UK [e] Department of Haematology, Oxford University Hospitals NHS Foundation Trust, Oxford, UK [f] Department of Medicine, Oxford University Hospitals NHS Foundation Trust, Oxford, UK [g] Adult Intensive Care Unit, Oxford University Hospitals NHS Foundation Trust, Oxford, UK [h] Clinical Microbiology, Department of Pathology, University of Virginia Health System, Charlottesville, Virginia, USA [i] Division of Infectious Diseases and International Health, Department of Medicine, University of Virginia Health System, Charlottesville, Virginia, USA [j]

ABSTRACT *Escherichia coli* and *Klebsiella* spp. are important human pathogens that cause a wide spectrum of clinical disease. In healthcare settings, sinks and other wastewater sites have been shown to be reservoirs of antimicrobial-resistant *E. coli* and *Klebsiella* spp., particularly in the context of outbreaks of resistant strains amongst patients. Without focusing exclusively on resistance markers or a clinical outbreak, we demonstrate that many hospital sink drains are abundantly and persistently colonised with diverse populations of *E. coli*, *Klebsiella pneumoniae* and *Klebsiella oxytoca*, including both antimicrobial-resistant and susceptible strains. Using whole genome sequencing (WGS) of 439 isolates, we show that environmental bacterial populations are largely structured by ward and sink, with only a handful of lineages, such as *E. coli* ST635, being widely distributed, suggesting different prevailing ecologies which may vary as a result of different inputs and selection pressures. WGS of 46 contemporaneous patient isolates identified one (2%; 95% CI 0.05-11%) *E. coli* urine infection-associated isolate with high similarity to a prior sink isolate, suggesting that sinks may contribute to up to 10% of infections caused by these organisms in patients on the ward over the same timeframe. Using metagenomics from 20 sink-timepoints, we show that sinks also harbour many clinically relevant antimicrobial resistance genes including bla_{CTX-M}, bla_{SHV} and *mcr*, and may act as niches for the exchange and amplification of these genes. Our study reinforces the potential role of sinks in contributing to Enterobacterales infection and antimicrobial resistance in hospital patients, something that could be amenable to intervention.

IMPORTANCE *Escherichia coli* and *Klebsiella* spp. cause a wide range of bacterial infections, including bloodstream, urine and lung infections. Previous studies have shown that sink drains in hospitals may be part of transmission chains in outbreaks of antimicrobial-resistant *E. coli* and *Klebsiella* spp., leading to colonisation and clinical disease in patients. We show that even in non-outbreak settings, contamination of sink drains by these bacteria is common across hospital wards, and that many antimicrobial resistance genes can be found and potentially exchanged in these sink drain sites.

46 Our findings demonstrate that the colonisation of handwashing sink drains by these
47 bacteria in hospitals is likely contributing to some infections in patients, and that
48 additional work is needed to further quantify this risk, and to consider appropriate
49 mitigating interventions.

50 **KEYWORDS:** enterobacterales, antimicrobial resistance, wastewater, resistome

51 **INTRODUCTION**

52 Infections caused by Enterobacterales, including *Escherichia coli* and *Klebsiella* spp.,
53 are major causes of global morbidity, and particular antimicrobial-resistant strains
54 (namely extended-spectrum beta-lactamase and carbapenemase producers) have been
55 listed as critical priority pathogens for mitigation by the WHO. In the UK, year-on-year
56 increases have been observed in the number of *E. coli* and *Klebsiella* spp. bloodstream
57 infections (1), for reasons which remain unclear. As well as causing invasive disease,
58 these organisms are capable of colonising a wide range of animal and environmental
59 niches, and are frequently carried in the human gastrointestinal tract (2). As such, they
60 are also commonly found in human wastewater, and in wastewater-associated sites
61 such as sewers and water treatment infrastructure (3).

62 A significant proportion of Enterobacterales infections are healthcare-associated,
63 prompting the UK government to introduce a target in 2016 to halve the number of
64 healthcare-associated Gram-negative bloodstream infections by 2021 (4). Wastewater
65 sites in hospitals have been highlighted as reservoirs of drug-resistant Enterobacterales,
66 with several studies reporting that ongoing transmission and outbreaks of human dis-
67 ease are associated with the contamination of, for example, sinks, by these organisms
68 (5, 6). More recently, several studies have shown reductions in colonisation and/or
69 invasive infection with Enterobacterales and other Gram-negative bacilli following the
70 introduction of strategies to remove sinks and mitigate possible contamination from
71 wastewater sources in patient rooms (7, 8). Most of these studies however focus on
72 the sampling and control of antimicrobial-resistant strains, often representing a more
73 immediate clinical problem in an outbreak setting, rather than on the possibility that
74 these sites may represent part of the wider endemic transmission network of both
75 susceptible and resistant strains causing infection in patients.

76 Whole genome sequencing of bacterial isolates is increasingly used as the most
77 robust, high-resolution approach to characterising relatedness between strains, and
78 hence determining likely transmission (9). However, the diversity of complex, polymicro-
79 bial environmental reservoirs is incompletely captured by sequencing small numbers
80 of isolates, and this breadth of diversity can be more fully captured by using a metage-
81 nomic approach, which characterises the genetic complement of a whole sample (10).
82 Combining both approaches has been shown to improve our understanding of species
83 and antimicrobial resistance (AMR) gene diversity within environmental, wastewater
84 and river samples (11, 12) and of transmission in a sink-associated outbreak of *Sphin-
85 gomonas koreensis* (also a Gram-negative bacillus) in the NIH Clinical Centre in the US
86 (13).

87 In order to investigate the prevalence of contamination of healthcare sinks by
88 strains of *E. coli* and *Klebsiella* spp., including those resistant to third-generation
89 cephalosporins and carbapenems, we sampled all sink sites using p-trap (U-bend)
90 aspirates across several wards and timepoints in a single UK hospital in 2017. We
91 used a combination of whole genome sequencing of cultured isolates from all sink
92 samples and metagenomic sequencing of a subset of sink samples to facilitate a high-

93 resolution assessment of the genetic diversity present in these niches. To determine
94 whether sinks were a reservoir of Enterobacterales strains causing infection in patients
95 over similar timeframes, we simultaneously retrieved relevant isolates from culture-
96 positive specimens taken from patients admitted to the same ward locations, and used
97 genomics to identify the degree of genetic relatedness.

RESULTS

99 **Diverse, often antimicrobial-resistant Enterobacterales strains are frequent—and**
100 **often persistent—colonisers of hospital sink drains.** 439 Enterobacterales isolates
101 comprising *E. coli* (n=180), *K. oxytoca* (n=166) and *K. pneumoniae* (n=93) were cultured
102 and successfully sequenced at one or more timepoints from 12/20 (60%), 9/23 (43%)
103 and 16/23 (70%) sinks sampled four times over 12 weeks (March-May 2017) in general
104 medicine (GM), adult critical care (ACC) and acute admissions (AA) wards respectively
105 (97/264 [37%] sink-timepoints culture-positive overall; Figure 1). A further 30 isolates
106 of *E. coli* (n=13), *K. oxytoca* (n=13) and *K. pneumoniae* (n=4) were cultured from 11/59
107 (19%) sinks in a haematology ward, sampled at a single timepoint only during this
108 period (Figure S1). See Table S1 for surveyed sink descriptions. Species distributions
109 (by culture) were relatively even across the general medicine ward, while the adult
110 critical care unit was enriched for *E. coli*, and the acute admissions ward was depleted
111 in *K. pneumoniae* (Table S2).

112 Analysis of whole genome sequences from cultured Enterobacterales revealed
113 widespread and sustained colonisation of sinks by multiple sequence types (STs) of
114 these species (Figure 1, Figure S1) In total, 8 known and 4 novel *E. coli* STs were repre-
115 sented, of which STs 635 (n=109, 61%), 401 (n=25, 14%) and 472 (n=18, 10%) accounted
116 for 84% of sequenced isolates (152/180). *Klebsiella* spp. STs were more varied: 15
117 known and 6 novel *K. oxytoca* STs were represented, of which the most frequent was
118 ST177 (33/166, 20%), while there were 18 known and 1 novel *K. pneumoniae* STs, the
119 most frequent being ST872 (24/93, 26%).

120 Across all locations and sink-timepoints, sequenced isolates comprised 20, 50 and
121 26 distinct strains (defined as differing by ≤100 recombination-adjusted core SNPs;
122 see Materials and Methods) of *E. coli*, *K. oxytoca* and *K. pneumoniae* respectively (Figure
123 1, Table 1). Positive sinks cultured up to three of these distinct strains per species
124 at any timepoint (Figure 1, Figure S1), reflecting significant diversity within species
125 in sink niches. Of the 37 longitudinally sampled culture-positive sinks from which
126 sequences were obtained, 31 (84%) grew isolates belonging to the same strain across
127 multiple timepoints, highlighting persistent background colonisation illustrated for
128 *E. coli* and *Klebsiella* spp. in respective figures 1 and S5. Isolates resistant to third-
129 generation cephalosporins were cultured at 16 sink-timepoints across 12 distinct sinks,
130 with resistant and susceptible cultures of the same genetic strain co-occurring in 11/16
131 (69%) sink-timepoints, suggestive of gain and/or loss of genes conferring cephalosporin
132 resistance in this setting. No carbapenem-resistant isolates were cultured.

133 **Enterobacterales can be highly abundant in sink drains, representing domi-**
134 **nant populations in some wards.** Deep metagenomic Illumina sequencing was per-
135 formed for 20 sink-timepoints on p-trap aspirates from seven sinks on the three wards
136 at two timepoints, and all four timepoints for a single sink unit in the adult critical
137 care ward (median 3.6m reads/sample; IQR: 3.3m-7.2m). The three most abundant
138 bacterial genera were *Klebsiella*, *Escherichia* and *Citrobacter*, all common healthcare-
139 associated pathogens (Figure 2). Sink drains in the general medicine ward were the
140 most abundantly colonised by Enterobacterales (Figure 2), to which more than 50%
141 of reads were assigned, and were markedly less diverse than those in adult critical

3

Constantinides et al.
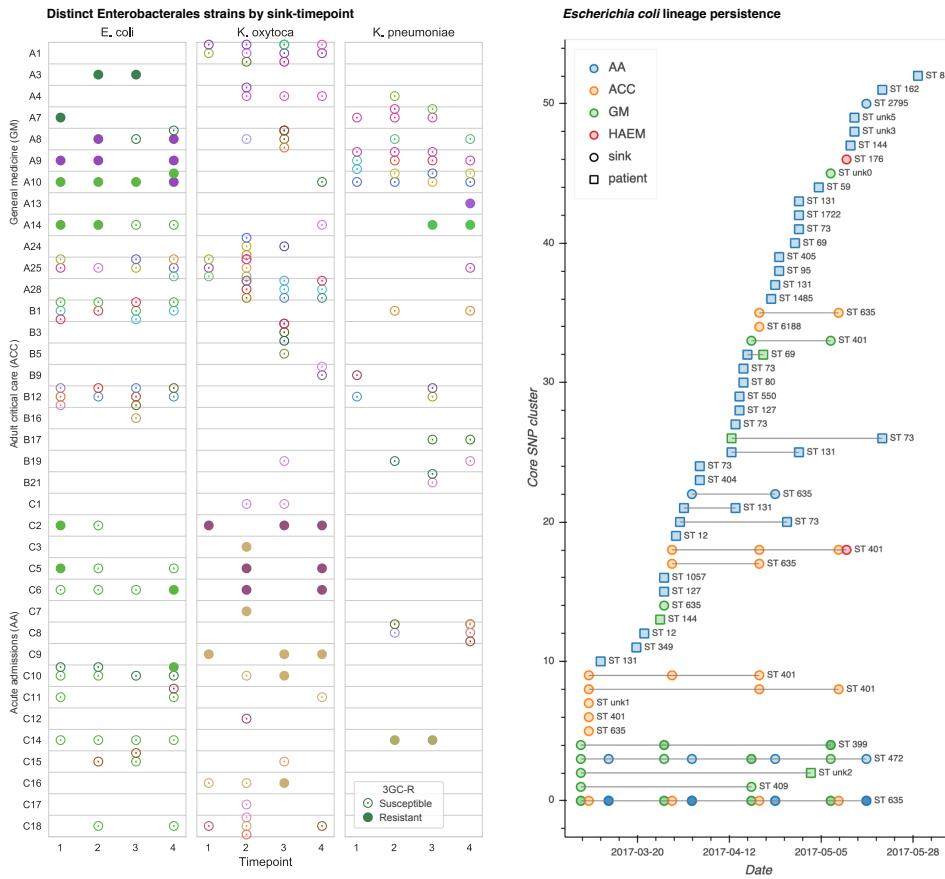
FIG 1 Cluster distribution and persistence. Left: strain-distinct cultured isolates of *E. coli*, *K. oxytoca* and *K. pneumoniae* from sink drain aspirates sampled over twelve weeks across three hospital wards. Different colours indicate distinct strains (defined by 100 SNP clusters), and cefpodoxime-resistant and/or selected ESBL-positive isolates are indicated by filled markers. Right: persistence of sink and contemporaneous patient *E. coli* strains throughout the sampling period.

A)		E.coli	K. oxytoca	K. pneumoniae	Total
Lineages with 1 isolate					
	Patient	28	1	3	32
	Sink	3	37	12	52
Lineages with >1 isolate					
Single timepoint	Same sink	5	6	3	14
	Different sinks; same ward	0	1	0	1
Multiple timepoints	Same sink	7	1	7	15
	Different sinks; same ward	1	4	4	9
	Different wards	3	1	0	4
Patients	Patient and sink (single timepoint)	1*	0	0	1
	Same patient	1	0	1	2
	Different patients	4	0	0	4
Total		53	51	30	134

lineage has 3 isolates; all taken from the same ward; 2 from the same sink at the same timepoint and 1 from a patient 2 months later.

B)		E.coli	K. oxytoca	K. pneumoniae	Total
Sink lineages with >1 isolate (excludes haematology ward)					
Single timepoint	Same sink	5	4	3	12
	Different sinks; same ward	0	0	0	0
Multiple timepoints	Same sink	8	1	7	16
	Different sinks; same ward	1	4	4	9
	Different wards	2	1	0	3
Total		16	10	14	40

TABLE 1 Spatiotemporal distribution of 100 core SNP lineages of cultured *E. coli* (n=53), *K. oxytoca* (n=51) and *K. pneumoniae* n=30) *A)* overall and *B)* occurring in >1 isolate in sinks on wards that were repeatedly sampled.

Constantinides et al.
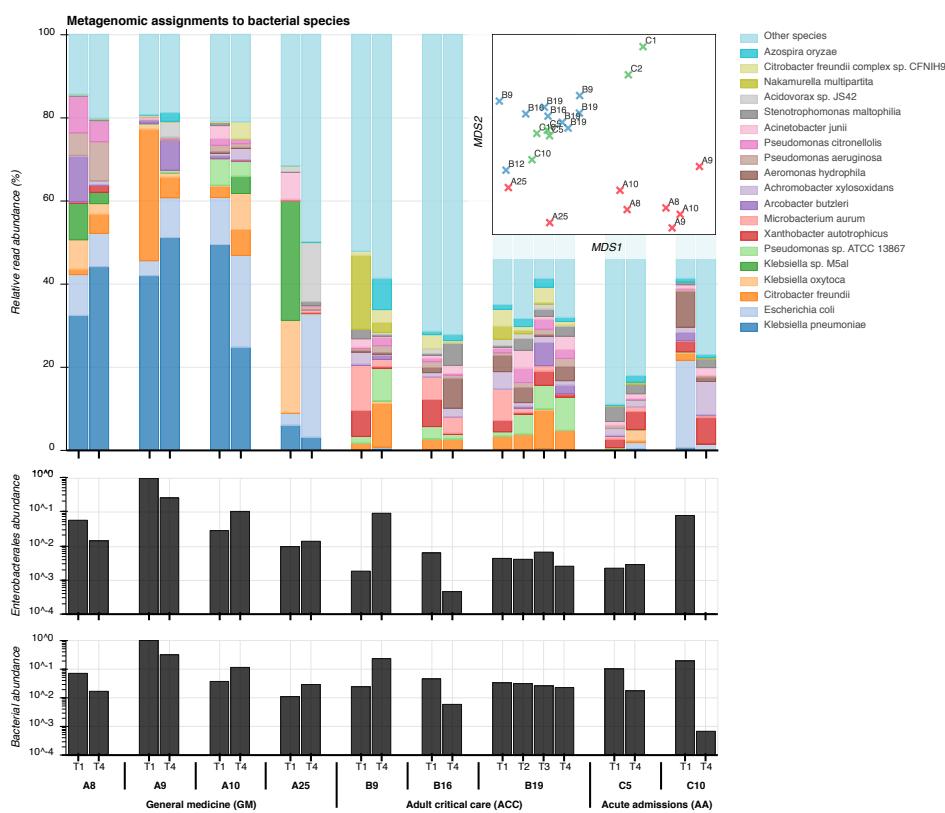
FIG 2 Taxonomic composition of sink microbiota from metagenomic sequencing. Top: relative abundance of the 20 most abundant bacterial species among sink drain aspirates (Kraken), inset with a corresponding multidimensional scaling (MDS) projection of pairwise distances between samples. Centre: spike-normalised relative abundance of species classifications at or below the order Enterobacterales among sink-timepoints. Bottom: spike-normalised relative abundance of Kraken classifications at or below the superkingdom Bacteria.

care and acute admissions wards, which had a dominance of *Klebsiella* spp., mirroring the culture results. 90% of species-level classifications in the sink-timepoints from the general medicine ward came from a median of just 21 bacterial species, compared with medians of 310 and 450 species in the adult critical care and acute admissions wards respectively. Microbial composition varied markedly between sampling timepoints for individual sinks, but sinks within wards exhibited more similar taxonomic profiles than those between wards (Figure 2), suggesting distinct ward-based wastewater ecologies. Total metagenomic sequence content was hierarchically structured by ward and by sink (Figure S2). Staff room sink A25 exhibited distinctive taxonomic and *k*-mer profiles from patient room sinks in the general medicine ward (Figure 2, Figure S2).

Sinks with high metagenomic abundance of the three Enterobacterales species reliably yielded corresponding cultures. The area beneath the receiver operating characteristic (ROC) curve for culture-based detection of these species was 0.93 (Figure S3). When the relative metagenomic abundance of a species was above 0.1%, 1% and 10%, one or more cultures of the same organism were obtained in 58% (18/31), 76% (16/21) and 89% (8/9) of sinks respectively. Conversely, culture detection therefore failed in 42% (13/31), 24% (5/21) and 11% (1/9) of cases where an Enterobacterales species was present at or above respective thresholds of 0.1%, 1% and 10% metagenomic abundance. A single sink-timepoint (first sample from A8; general medicine) failed

161 to culture any Enterobacterales, but yielded 4%, 5% and 16% relative metagenomic
162 abundances for the three study species. Thus metagenomic sequencing suggests that
163 persistence may be even more widespread than indicated by culture alone.

164 **Most environmental Enterobacterales appear to cluster within specific sinks**
165 **and wards, except for *E. coli* ST635, which is widely distributed.** The 96 strains
166 found in sinks exhibited structure at the ward and sink level. Four strains were found
167 in multiple wards (predominantly *E. coli* ST 635, but also STs 472, 401 and *K. oxytoca*
168 ST 146). 92 (96%) strains were only ever found in a single ward, and of the 44 strains
169 cultured twice or more, only 14 (32%) were cultured from different sinks (Table 1A).
170 Further, of the 40 strains cultured twice or more on wards which were repeatedly
171 sampled (i.e. excluding the haematology ward which was only sampled once), 12
172 (30%) were only seen at the same sink-timepoint, 16 (40%) were seen in the same
173 sink at different timepoints, and 12 (30%) in different sinks at different timepoints
174 (1B). This structure was reflected in the recombination-adjusted core genome species
175 phylogenies (Figure 3). The main exception to ward and sink-based clustering was *E.*
176 *coli* ST635, which comprised more than half of isolates sequenced from sinks, and was
177 found in 13/20 (65%) *E. coli*-positive sinks.

178 However, there was sink-level clustering even within *E. coli* ST635, more clearly
179 shown in neighbour-joining trees constructed from pairwise read-based MASH dis-
180 tances, representing both core and accessory genomic content (see *E. coli* ST635 zoom
181 in Figure 3; colours indicate distinct sinks). Although pairwise correlations between core
182 and accessory genomic distances were high (Table S3), incorporating accessory content
183 yielded additional resolution beyond core SNP distances (Figure S4). Permutational
184 analysis of variance (PERMANOVA) using pairwise core SNP and read-based MASH
185 distances supported significant grouping of isolates from all three species by ward and
186 by sink (P<0.001), most conclusively for *K. pneumoniae* (Table S4).

187 **Patient *E. coli* isolates were more diverse than those found in sinks, and**
188 **included isolates from known 'high-risk' clinical lineages.** From March to May 2017,
189 1384 relevant clinical samples from 719 patients were submitted to the microbiology
190 laboratory for processing (AA n=779, ACC n=365, GM n=240), of which 397/1384 (29%)
191 were positive for microorganisms, and 107 were culture-positive for one of the study
192 organisms (*E. coli* [n=96], *K. oxytoca* [n=2], *K. pneumoniae* [n=9]). 46/107 (43%) of these
193 isolates were retrieved for sequencing, including 19/22 isolates from bloodstream
194 infections, 3/6 from respiratory samples, 21/73 from urine samples, and 0/10 other
195 samples.

196 Among 39 sequenced *E. coli* patient isolates, 21 STs were represented, including
197 known high-risk lineages (23/39 (59%) isolates) not seen in sinks: namely ST73 (n=8),
198 ST131 (n=7), ST69 (n=3), ST12 (n=2), ST127 (n=2), and ST95 (n=1). The single sequenced
199 *K. oxytoca* isolate was ST36, and the six *K. pneumoniae* isolates came from four STs,
200 including two high-risk lineages, ST25 and ST29. Across the three species, there were
201 34, 1 and 4 distinct lineages, respectively (Table 1).

202 **Genetic similarity of sequenced patient and environmental isolates.** As well
203 as being diverse, the 39 clinical *E. coli* isolates were phylogenetically distinct from most
204 sink isolates, which largely came from just four sequence types (ST635, ST401, ST472
205 and ST399). The exception was an *E. coli* isolated from urine taken on the general
206 medicine ward, which was 17 and 19 core SNPs from two isolates from sink A25 in the
207 same ward, sampled 58 days prior to the clinical sample (Figure 1; right; cluster 4). A
208 read-based MASH distance of 7×10^{-6} between this pair of isolates indicated very high
209 total genomic (chromosome+accessory) similarity. A records search for admissions
210 of this patient prior to commencing sink sampling revealed four inpatient admissions

Constantinides et al.
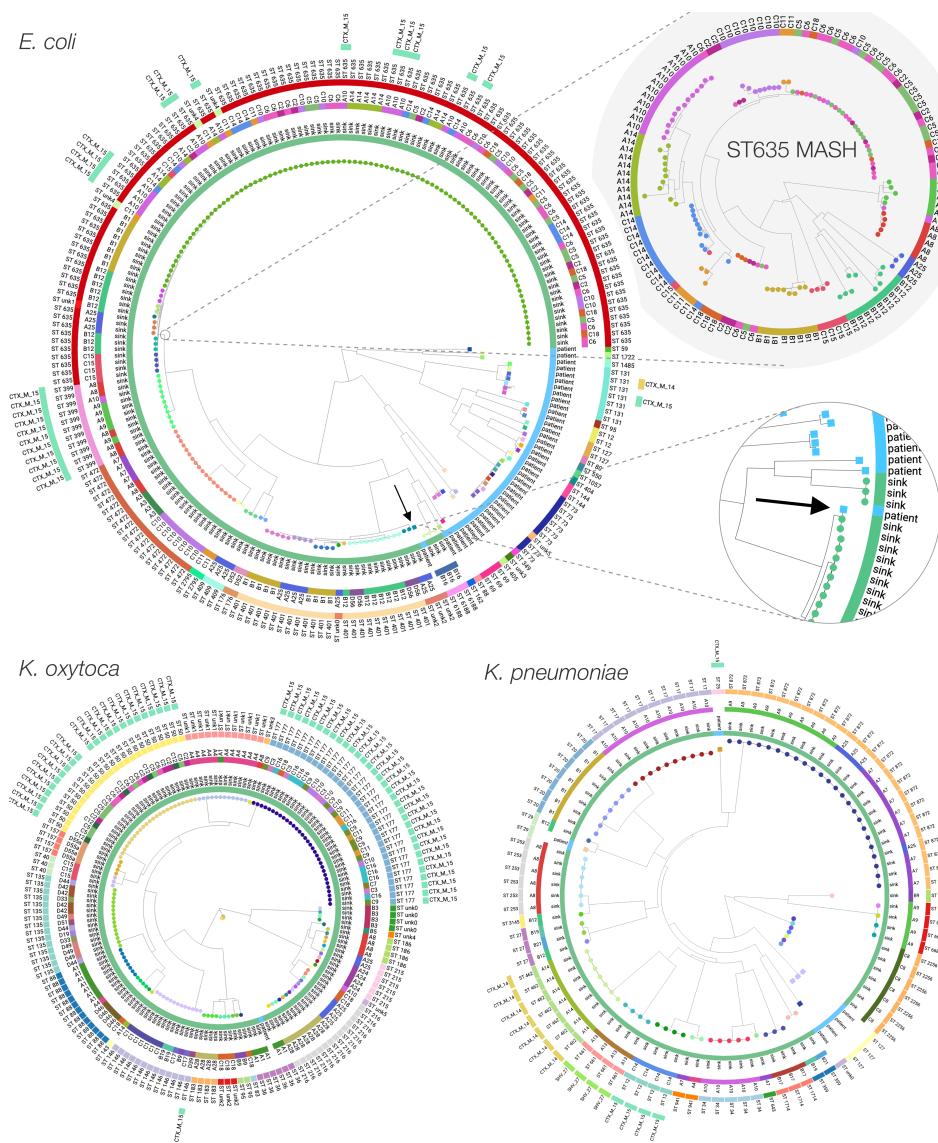
FIG 3 Maximum likelihood phylogenies of *E. coli*, *K. oxytoca* and *K. pneumoniae* cultured from sink drain aspirates sampled over twelve weeks across three wards, with two zooms corresponding to an *E. coli* ST635 neighbour-joining MASH subtree whose tips are coloured by sink, and genetic overlap between a sink culture and a urine culture from a patient with ward contact during the study. Tip colours indicate strains, with rings inside-to-out denoting: patient/sink, sink designation, sequence type, and ESBL genotype.

211 (of 0 (day case), 1, 2 and 5 nights' duration), of which 3 included time on the acute
212 admissions ward. There were no prior admissions onto the general medicine ward (in
213 which their positive urine specimen was taken), although their eleven-night spell on the
214 general medicine ward commenced with a seven-hour episode in acute admissions.
215 The positive clinical specimen was taken ten days after the patient's admission onto the
216 general medicine ward, indicating a large duration of exposure to a ward environment
217 shown to be harbouring a very similar strain of *E. coli* to the patient's urine culture.
218 The next most closely related *E. coli* clinical isolate was 3,688 core SNPs from its
219 nearest sink neighbour (read-based MASH distance 0.009), reflecting the otherwise
220 large evolutionary distances separating the cultured clinical and environmental *E. coli*
221 (Figure 3).

222 Unlike *E. coli*, the small numbers of *Klebsiella* spp. patient isolates were not phylo-
223 genetically segregated from environmental isolates, but the closest patient and sink
224 isolates differed by 2,558 core SNPs, indicating a lack of observed overlap over these
225 timeframes.

226 **Antimicrobial resistance genes are prevalent and spatially structured in sinks.**
227 The presence of 571 clustered CARD antimicrobial resistance genes in cultured iso-
228 lates was supported by $\geq$75% exactly matching read coverage reported by ResPipe
229 (Figure 4). Among these were known transmissible genes of clinical concern includ-
230 ing beta lactamases (e.g. bla_{TEM}, bla_{CTX-M}, bla_{SHV}), aminoglycoside resistance genes
231 (*aac(3)*, *aac(6)* families) and quinolone resistance genes (*qnr* family). Some of these,
232 including cmlA and qacH, were widely seen in sink metagenomes but less frequently
233 in cultured isolates, consistent with a background resistance reservoir that may pose
234 a risk in different populations to those cultured (of either same or different species).
235 Spatial structure was evident among both cultured isolates and metagenomes, al-
236 though resistance repertoires of isolates frequently clustered across ward boundaries,
237 in keeping with findings of our prior core genome analysis. Resistance genes detected
238 in cultured sink isolates were also abundant within sink metagenomes at one or more
239 timepoints. Sink drain metagenomes yielded 673 CARD genes exceeding 75% coverage,
240 and after clustering large gene families represented by many similar sequences (see
241 methods for detailed description), only five genes abundant in one or more cultured
242 isolates were not detected in at least one metagenome. Notably, these five genes
243 (*gadW*, *len-26*, *tet(B)*, *mgrA* and *sat-2*) were all seen in isolates from sinks not subject to
244 metagenomic sequencing, showing that resistance genes cultured from sink drains
245 were highly contained in corresponding metagenomes.

246 Third generation cephalosporin-resistant phenotypes in Enterobacterales sink iso-
247 lates could be explained by the presence of major extended-spectrum beta-lactamase
248 (ESBL) genes blaSHV-27, $bla_{CTX-M-14}$ and $bla_{CTX-M-15}$, detected by ARIBA/CARD in 4, 8 and
249 87 isolates respectively. $bla_{CTX-M-15}$ was identified in two distinct strains of *E. coli* ST635
250 and ST399, restricted to three bay sinks (A8, A9, A10) in three adjacent rooms of the
251 general medicine ward. $bla_{CTX-M-15}$-positive *K. oxytoca* ST50 and ST177 were identified in
252 10 sinks (C2-3, C5-6, C7, C9-12, C16) on the acute admissions ward. bla_{SHV-27}, $bla_{CTX-M-14}$
253 and $bla_{CTX-M-15}$ were observed in *K. pneumoniae* from sinks A13, A14 (general medicine)
254 and C14 (acute admissions) respectively; one *K. pneumoniae* patient isolate was also
255 $bla_{CTX-M-15}$-positive. These findings suggest sink-associated isolates, such as *E. coli*, may
256 represent reservoirs of clinically relevant resistance genes.

257 Surprisingly, the colistin resistance gene *mcr-4* was detected in the metagenomes—yet
258 not cultured genomes—of three sinks in adjacent bays of the general medicine ward
259 (A8-10). Assembly of the sink A10 metagenome generated a 5.4kbp plasmid sequence
260 containing an *mcr-4* gene with 98.8% overall identity at 94% query coverage to an

9

Constantinides et al.
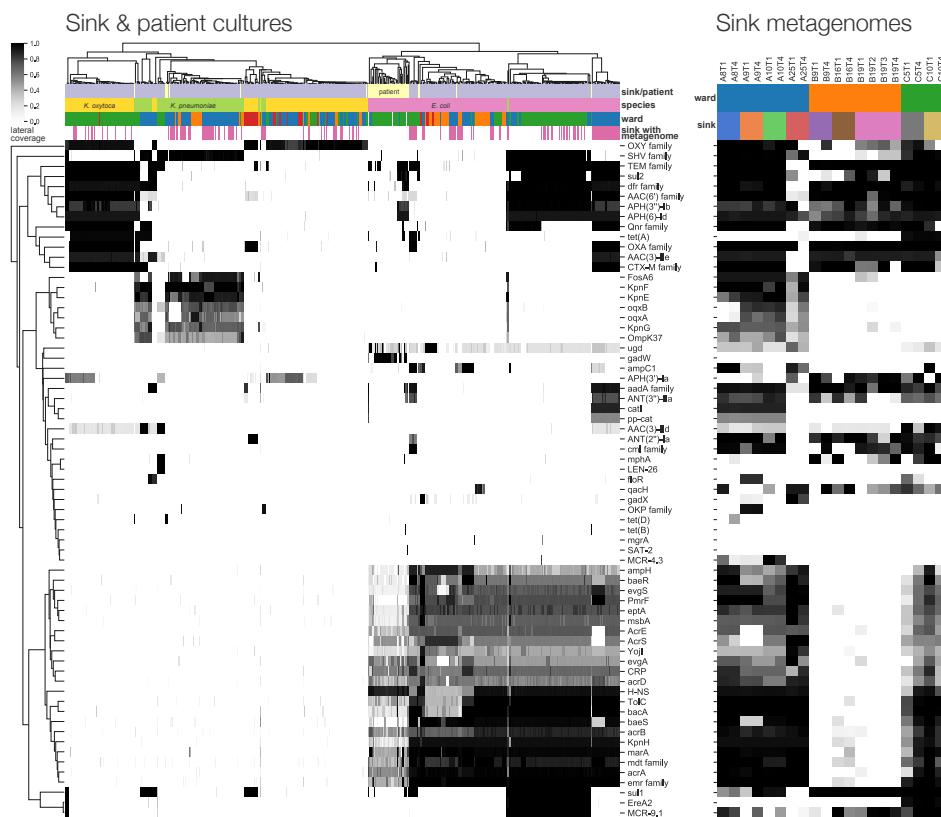
FIG 4 Antimicrobial resistance gene content of cultured isolates and sink drain metagenomes. Left: lateral coverage of ResPipe/CARD genes within sink drain and clinical isolates. Displayed genes attained ≥75% lateral coverage in one or more isolates. Right: Corresponding lateral coverage of the same genes in sink drain aspirate metagenomes.

Multi-omic surveillance of *Escherichia coli* and *Klebsiella* spp. in hospital sink drains and patients

261 8.7kbp pMCR-4.2 plasmid previously reported in pigs from Italy, Spain and Belgium (14).
262 This *mcr* variant has been previously reported in European Acinetobacter, Enterobacter,
263 Salmonella, and Escherichia spp. but not to our knowledge in the United Kingdom.
264 Screening all metagenomes for assembled *mcr-4* produced alignments in two sinks
265 on the ward (A8, A9) across a total of six sink-timepoints, with coverage and abun-
266 dance suggesting low and declining prevalence of this gene over time (Table S5). An
267 *mcr*-positive *E. coli* (reported as *mcr-4.3*) was cultured from sink C5 (acute admissions;
268 second timepoint) according to both ARIBA/CARD and ResPipe/CARD. Metagenomic
269 sequencing was performed for the first and the fourth but not the second timepoint
270 aspirate for this particular sink. Another *mcr* gene, *mcr-9* was more widespread, and
271 detected with complete coverage in 77 cultured isolates across 11 distinct sinks, pre-
272 dominantly but not exclusively in *E. coli* (73/77 occurrences).

273 **Metagenomic screening suggests that clinical isolates may be more widely**
274 **present in the environmental reservoir than observed from culture-based com-**
275 **parisons.** Sink metagenomes were individually screened for *k*-mer containment of *i)*
276 strain-representative sink isolate genome assemblies, *ii)* strain-representative patient
277 isolate genome assemblies and *iii)* core genomes of selected control organisms, in-
278 cluding five clinical core genomes each from pathogenic strains of *E. coli* and *Klebsiella*
279 spp. from Bush et al. (15), together with NCBI canonical species references for several
280 pathogens expected to be absent from sink drain microbiota (Figure 5; left). This
281 demonstrated similar sink, ward, and temporal structure to that of culture, particu-
282 larly underlying similarities in the microbiota of nearby sinks, as well as flux between
283 sampling timepoints. Strain-representative sink culture assemblies from the same sink
284 and timepoint as the screened sink metagenome were the best contained, sharing the
285 most *k*-mer hashes. Assembled isolates originating from the same sink but at a differ-
286 ent timepoint to the screened metagenome shared significantly fewer *k*-mer hashes
287 (P=0.013) than same sink/same timepoint comparisons. The containment of strain-
288 representative assemblies from different sinks in the same ward as the metagenome
289 was significantly lower still (P<0.0001), and so in turn were the remaining comparisons
290 of cultures grown from different sinks in different wards to the screened metagenome
291 (P<0.0001).

292 Among control genomes, *k*-mer hashes shared between sink metagenomes and
293 the core genomes of *Neisseria gonorrhoeae*, *Staphylococcus aureus*, *Clostridioides difficile*,
294 *Enterococcus faecalis* and *Vibrio vulnificus* did not exceed 3%. Reference genomes of *E.*
295 *coli*, *K. pneumoniae* and *Pseudomonas aeruginosa* were abundant and highly contained
296 by many metagenomes, but none exceeded 90% shared *k*-mer hashes (Figure 5; right).

297 In contrast, screening for strain-representative patient assemblies in sink metagenomes
298 revealed significantly greater similarity (P<0.0001) between patient and environmen-
299 tal Enterobacterales strains collected from the same ward than from different wards
300 (Figure 5; centre), supporting genetic overlap between clinical isolates from patients
301 and uncultured isolates in sink niches in a given ward setting. Indeed, the only strain-
302 representative patient isolate with greater than 90% sink metagenome *k*-mer contain-
303 ment was the *E. coli* urine culture described in the aforementioned case of sink-patient
304 overlap, of which 99.5% and 93.0% of *k*-mers were contained within the respective
305 A25T1 and A25T4 sink metagenomes (Figure 5; centre; red markers).

DISCUSSION

307 In this study, we have demonstrated that hospital sink drains are widely—and in many
308 cases abundantly—contaminated with key Enterobacterales species causing healthcare-
309 associated infections, and are potential reservoirs of multiple resistance genes en-

11

Constantinides et al.
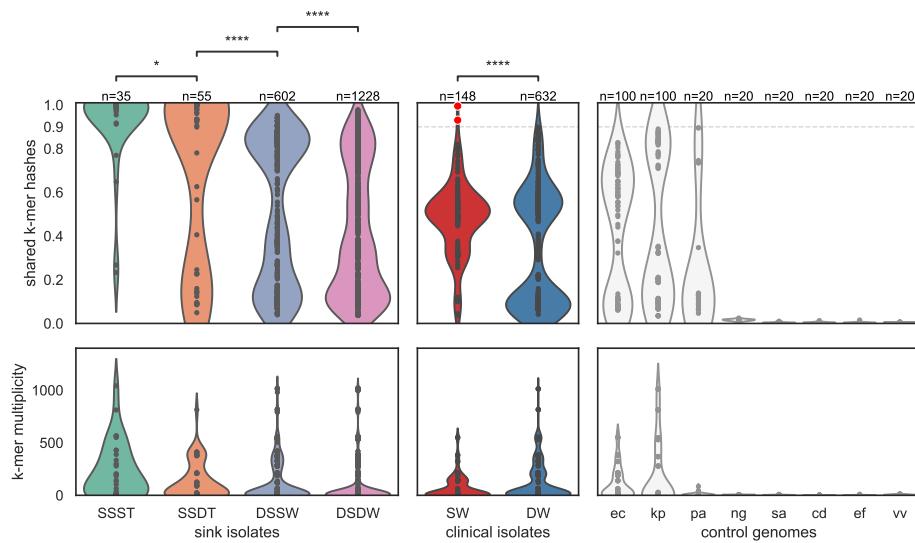
FIG 5 **Metagenomic containment of sink (left) and patient (centre) cultured strain-representative genome assemblies, and control genomes (right). Shared *k*-mer hashes and median *k*-mer multiplicity values are as reported by MASH Screen. SSST=same sink and same timepoint; SSDT=same sink at different timepoints (shared hashes Mann-Whitney U P=0.013 vs. SSST); DSSW=different sinks of the same ward (P<0.0001 vs. SSDT); DSDW=sinks on a different ward (P<0.0001 vs. DSSW). SW=strain-representative assemblies of clinical isolates in the same ward; DW=strain-representative assemblies of clinical isolates from a different ward (P<0.0001 vs. SW). Control genomes comprised *E. coli*, *K. pneumoniae P. aeruginosa*, *N. gonorrhoeae*, *S. aureus*, *C. difficile*, *E. faecalis*, and *V. vulnificus*, shown abbreviated with binomial initials. A case of within-ward sink-patient overlap is highlighted with red markers, corresponding to high strain containment in the metagenomes of sink A25 timepoints 1 and 4.**

coding resistance to important clinical antimicrobials. Populations of antimicrobial-resistant and susceptible *E. coli* and *Klebsiella* spp. may be persistent colonisers of sinks, and different wards may have markedly different sink ecosystems, perhaps reflecting different and potentially modifiable infrastructures, selection pressures, and contributing sources. Ward and sink-level genetic structure was most evident within the accessory genome, and observed repertoires of transmissible resistance genes often transcended species boundaries, instead clustering more tightly by sink unit. Characterising these highly diverse reservoirs is difficult, and we have shown that combination approaches utilising metagenomics and sequencing of cultured isolates are complementary in understanding the diversity of species, strains, and the resistance genes present within these niches. For example, metagenomics highlighted several cases of abundant *mcr-4* in sink drain aspirates from which cultured Enterobacterales isolates did not carry the gene.

Colonisation patterns of sink niches differed markedly between the two genera investigated. *E. coli* strains have evolved to colonize and adapt to multiple niches, including some which have adopted pathogenic lifestyles, and appear to have different distributions in humans, domesticated and wild animals, and the environment. There is however no absolute correlation between phylogenetic lineage and any given niche, and overlaps are observed. Interestingly, in our study, more than half of the *E. coli* sink isolates cultured were ST635, which has been recently described as a highly adapted, resistance- and virulence gene-enriched wastewater-associated strain thought to be globally distributed, but is also found in humans, animals and other environments (16). Of note, it has been observed in association with several clinically relevant transmissible resistance genes, including ESBLs, carbapenemases, and rRNA methylases, and was one of only two *E. coli* STs in our study that harboured an ESBL (bla$_{CTX-M-15}$). We observed presence/absence of bla$_{CTX-M-15}$ across closely related ST635 isolates, suggesting that this gene may be frequently lost/gained in sinks. Also notable in the context of ST635 was the ability of read-based *k*-mer composition to resolve fine-grained structure between the populations of different sinks, beyond that observed in the core-only SNP phylogeny. Other common *E. coli* sink lineages were ST399 and ST472, which to date have predominantly been seen in humans/animals, rather than the environment.

The phylogenetic distribution of sink isolates of *K. pneumoniae* appeared to mirror that seen in a global collection of isolates, providing little evidence that a particular lineage was predominating in, or particularly adapted to, the wastewater environment. Studies of the population structure of unselected *K. oxytoca* are limited, but again we observed a diverse population amongst sink isolates, with a deep branch separating two distinct groups as previously described. Interestingly, two *K. oxytoca* strains associated with bla$_{CTX-M-15}$ were widely distributed amongst sinks in the acute admissions ward; outbreaks of ESBL- and carbapenem-associated *K. oxytoca* in association with contaminated handwashing sinks have been described in other settings (17).

Genomic overlap with sink isolates was identified in 1/46 (2%; 95% CI: 0.05-11%) of all sequenced isolates causing clinical infections over the same timeframe, with a temporal association consistent with acquisition from a sink source (i.e. sink isolate observed first), and following ten days of patient exposure to a ward environment wherein the overlapping strain was previously cultured. We may have significantly underestimated the degree of overlap between these two compartments for several reasons. Firstly, we have shown the diversity in sink niches is substantial, and with a culture-based approach agnostic to any selective marker, even sequencing 444 isolates from 48 sinks will have limited ability to capture the underlying diversity for complete

comparison of sink-patient pairs at the isolate-level. Supporting this, screening the metagenomes of a subset of 20 sinks using patient isolates suggests that overlap between these reservoirs may be more common than observed at the isolate-level. Second, clinical isolates represent the tip of the iceberg of any transmission chain, with the majority of transmission events likely occurring between gastrointestinal tract (asymptomatic carriage) and the wastewater environment. Nonetheless, in the context of understanding how sinks may be contributing to infection caused by *Klebsiella* spp. and *E. coli*, focusing on clinical isolates seems appropriate. Third, the interval between sampling dates for our observed patient-sink isolate-pair was 58 days, suggesting that the timeframe between acquisition from the environment and infection may be long, and may not be adequately captured with a study timeframe spanning three months.

In addition, a major study limitation is the fact that only 46/107 patient isolates could be successfully retrieved (due to the high turnover of samples in our high-volume service laboratory), and *Klebsiella* spp. cultures were especially limited. The risks of transmission and possibly sink-associated infection could be more clearly defined by more extensive sampling over a greater timeframe, and thorough investigation into the exchange of resistance-associated mobile genetic elements, but would require a considerable increment in resource. Characterising microbial diversity present on sink strainers would also be of benefit, as the risks of droplet-mediated dispersal from sink drains have been shown to be most significant when the sink drain is located immediately below the tap, and if the organisms migrate from the sink trap onto the strainer (18, 19). However, given the different sink structures across wards, the p-trap was the only site which could be consistently sampled (since ACC had horizontally draining sinks without strainers). Characterising factors that might be associated with greater predominance of Enterobacterales and drug-resistant Enterobacterales, such as sink usage, ward-level antimicrobial usage, and patient populations, would also be of interest.

In conclusion, without conditioning on the presence of resistance markers, we have demonstrated that colonisation of ward sink drains with diverse and abundant populations of Enterobacterales, including drug-resistant strains, is common and persistent. The evidence linking contaminated, unmitigated wastewater reservoirs (including sink drains) in healthcare settings with outbreaks of colonisation/disease with drug-resistant Gram-negative bacilli in patients seems clear (5, 20), but no study to our knowledge has focused on the potential risk posed by Enterobacterales in sinks in general. Screening of sinks is not carried out in the absence of observed outbreaks, making it difficult to quantify wider patient-associated risk from the studies available. We demonstrate that contaminated sinks may be contributing to a proportion of healthcare-associated infections caused by Enterobacterales, and further work to investigate how to reduce the risk posed by this hospital environmental reservoir is warranted.

MATERIALS AND METHODS

Ward-based sink sampling. We sampled three units (acute admissions [AA], adult critical care [ACC], adult general medicine [female only] [GM]) within a single hospital (John Radcliffe Hospital, Oxford, UK) four times on rotation every three weeks over three months, March-May 2017. Units were chosen to capture different patient populations, admission turnaround times and wastewater plumbing infrastructure. The haematology ward (on a separate hospital site) was also sampled on a single day (12th May 2017) subsequent to a small cluster of patient cases of bla$_{OXA-48}$ carbapenemase-associated Enterobacterales bloodstream infections [described previously (21). Ward

409   and sink/wastewater layouts were obtained from estates, and each sink/drain site was
410   assigned a unique site identifier (Table S1).

411   On each day of sampling, autoclaved tubing cut to 10 inches was used to aspi-
412   rate from sink p-traps via a sterile 50ml syringe. Up to 50mls of fluid was aspirated
413   where possible. 100µL of 10-fold dilutions (10^{-2}, 10^{-3}, 10^{-4}) of each sink p-trap aspirate
414   were plated onto CHROMagar Orientation media (Becton Dickinson, Franklin Lakes, NJ,
415   USA), with no disc, cefpodoxime (10µg), ertapenem (10µg) (Thermo Scientific Oxoid,
416   Basingstoke, UK) applied in a triangular fashion to each plate. Cultures were incu-
417   bated at 37°C for ~18hrs. Growth of Enterobacterales (presence/absence) and density
418   (sparse/dense/confluent) in all zones was recorded (i.e. no antibiotic, in the presence of
419   cefpodoxime, and in the presence of ertapenem). Up to four distinct colonies of each
420   of presumptive *E. coli* and *Klebsiella* spp. were sub-cultured on CHROMagar Orientation
421   to confirm purity and species identification. Species identification of sub-cultured
422   colonies was confirmed by MALDI-ToF (MALDI Biotyper, Bruker, Billerica, MA, USA).
423   Stocks of sub-cultured isolates were stored at -80°C in 400µl of nutrient broth + 10%
424   glycerol prior to DNA extraction for sequencing. Aspirates from sink p-traps were then
425   centrifuged at 4000 rpm for 10 minutes at 4°C, and supernatants removed; pellets
426   were stored at -80°C.

427   **Patient isolate sampling.** For AA, ACC, GM wards, a pseudo-anonymised, prospec-
428   tive feed was set-up to try and enable real-time capture of isolates from all samples
429   culture-positive for *E. coli*, *K. pneumoniae* and *K. oxytoca* from patients that had been
430   admitted to any of these wards during the study time period and were processed
431   routinely through the clinical microbiology laboratory in the John Radcliffe Hospital in
432   accordance with local standard operating procedures for clinical sample types, and
433   compliant with national standards for microbiology investigations (22). These typi-
434   cally involve selective culture steps and species identification using MALDI-ToF (MALDI
435   BioTyper, Bruker, Billerica, MA, USA).

436   Pseudo-anonymised extracts of all patient culture results and admission/discharge
437   data covering the study period were obtained after the study was finished through
438   the Infections in Oxfordshire Database (which has generic Research Ethics Committee,
439   Health Research Authority and Confidentiality Advisory Group approvals [14/SC/1069,
440   ECC5-017(A)/2009]) to enable an evaluation of *i)* baseline sampling denominators,
441   *ii)* the extent of relevant clinical isolate capture, and *iii)* the temporal and spatial
442   overlap of any genetically related sequenced isolates from patients and sequenced
443   isolates/metagenomes from sinks.

444   **Isolate sequencing and p-trap aspirate metagenomics.** All isolates confirmed
445   as *E. coli*, *K. pneumoniae* and *K. oxytoca* from patients and p-trap aspirates were ex-
446   tracted for sequencing using the QuickGene DNA extraction kit (Autogen, MA, USA) as
447   per the manufacturer's instruction, plus an additional mechanical lysis step prior to
448   chemical lysis (FastPrep, MP Biomedicals, CA, USA; 6m/s for two 40 second cycles).

449   For metagenomics, DNA was extracted from a subset of stored pellets (n=20)
450   using the MoBio PowerSoil DNA isolation kit (Qiagen, Hilden, Germany) as per the
451   manufacturer's instructions, and including a mechanical lysis step of two 40 second
452   cycles at 6m/s in lysing matrix E and final elution in buffer CDT-1 (Autogen, MA, USA).
453   45ng of Thermus thermophilus DNA (reference strain HB27, ATCC BAA-163 [DSMZ,
454   Germany]) was added to each sample in the PowerBead tube at the start of the
455   experiment, prior to the addition of solution C1 as an internal control and normalisation
456   marker (12). Sink aspirates were selected for metagenomics sequencing to enable
457   evaluation of *i)* microbiome differences within and between wards, *ii)* longitudinal
458   change in microbiota composition, and *iii)* whether culture-negative sinks harboured

459 the bacterial species being studies i.e. indicating limited sensitivity of culture-based
460 approaches.

461 Short read sequencing (single isolate and metagenomics) was performed on the
462 Illumina HiSeq 4000, pooling 192 isolate extracts and 6 metagenomes per lane, and
463 generating 150bp paired-end reads.

464 **Computational methods.** Cultured isolate informatics. Of the isolates sent for
465 sequencing, 439/446 (98%) sink and 46/46 (100%) patient isolates were successfully
466 sequenced and classified with Kraken/MiniKraken (23) as Enterobacterales, and used
467 for subsequent analysis. Isolate consensus sequences were constructed by read
468 mapping and consensus inference with respective *E. coli*, *K. oxytoca* and *K. pneumoniae*
469 reference genomes AE014075.1, NC_018106.1 and CP000647.1 using Snippy 4.4.0 (24).
470 Isolate genomes were assembled using Shovill 1.0.4 (25). Recombination-adjusted
471 phylogenetic reconstruction was performed using runListCompare 0.3.8 (26) wrapping
472 IQ-TREE 1.6.11 (27) and ClonalFrameML 1.12 (28). Final core genome alignments
473 included 218/219 *E. coli* isolates, 165/167 *K. oxytoca* isolates and 98/99 *K. pneumoniae*
474 isolates, all of which satisfied the runListCompare filtering criteria of perACGT_cutoff
475 ≥70%, varsite_keep ≥0.8 and seq_keep≥0.7. 100 SNP core genome clusters were
476 defined by single linkage clustering of runListCompare pairwise distance matrices.
477 Trees were midpoint rooted prior to visualisation. See supplementary data repository
478 for runListCompare configuration. Read-based MASH trees were constructed using
479 MASH 2.2.2 (29) and RapidNJ 2.3.2 (30) using 21mers, a sketch size of 10,000 and a
480 minimum abundance threshold of 10 *k*-mers. Assembly-based core and accessory
481 genome partitioning was performed using PopPUNK 1.1.7 (31). Resistance genotyping
482 and phenotype prediction in cultured isolates was performed using ResPipe and ARIBA
483 2.14.4 (32) with the CARD 3.0.3 database (33). Tree comparisons (tanglegrams) were
484 generated using the R package Dendextend 1.5.0 (34).

485 **Metagenome informatics.** Metagenomic sequences were analysed for taxonomic
486 and antimicrobial resistance gene presence using ResPipe (12) and Kraken2 (35) with
487 CARD database version 3.0.3. Large resistance gene families were clustered to facilitate
488 visualisation of resistance profiles (Figure 4) (methodology documented in supplemen-
489 tary data repository). A metagenomic assembly of the *mcr-4* gene was generated with
490 MEGAHIT 1.2.9 (36), to which reads were aligned with Minimap2 2.17-r941 (37) and
491 consensus inferred using Kindel (38). Metagenomic summary statistics were generated
492 using Pavian (39). Data analysis was performed with the SciPy ecosystem (40) and
493 JupyterLab (41). Matplotlib (42), Bokeh and Microreact (43) were used for visualisation.

494 **Data availability.** Raw sequencing data are available under NCBI SRA accessions
495 PRJNA604910 and PRJNA604975 (cultured isolates), and ENA project PRJEB36775
496 (metagenomes). A supplementary data repository containing metadata, phyloge-
497 nies, Jupyter notebooks, Microreact projects and Pavian reports is archived at https://
498 figshare.com/articles/Enterobacterales_colonisation_of_hospital_sink_drains/11860893

SUPPLEMENTARY MATERIAL

Table S1. Surveyed sink descriptions and *E. coli*/*Klebsiella* spp. culture results across timepoints.

Table S2. Cultured Enterobacterales species by ward.

Table S3. Pairwise Mantel correlation of different within-species distance matrices. These include recombination-adjusted core SNP phylogeny (reads-core-snp), read-based MASH distance (reads-mash) and PopPUNK estimates of core and accessory genomic distance from de novo assemblies (assemblies-core-mash, assemblies-accessory-mash).

Table S4. Permutational analysis of variance. Permutation tests for association of genetic structure with ward (n=3) and sink (n=18) for three species of sink drain Enterobacterales. Corresponding test results are shown for differential dispersion between groups (PERMDISP). Bold type indicates significant (p<0.05) group association under PERMANOVA in the absence of significant differential dispersion (PERMDISP).

Table S5. *mcr-4* coverage. Sequencing coverage and mean depth of the 1,626bp metagenome-assembled *mcr-4* gene from sink A10, to which metagenomic short reads mapped from three sinks (including A10) across six sink-timepoints within the general medicine ward.

Figure S1. Cultured strains observed on the Haematology ward. Different colours indicate distinct 100 core SNP strains, and cefpodoxime-resistant and/or ESBL gene-positive isolates are indicated by filled markers.

Figure S2. Spatial structure of sink metagenome *k*-mer composition. Left and centre: visualisation of 31mer pairwise MASH distances of total metagenome content using hierarchical clustering (left) and multidimensional scaling (centre). Right: comparison of within sink, within ward and between ward pairwise MASH distances.

Figure S3. Receiver operating characteristic (ROC) for detection of Enterobacterales by culture with varying metagenomic abundance.

Figure S4. Tanglegrams comparing recombination-corrected core phylogenies and read-based whole genome MASH + neighbour joining phylogenies for a) *E. coli*, b) *K. oxytoca* and c) *K. pneumoniae*. Topologically consistent subtrees are rendered with solid branches.

Figure S5. *Klebsiella* spp. lineage persistence in cultured sink drain aspirates and contemporaneous clinical isolates from patients with ward contact during the sampling period.

Constantinides et al.

REFERENCES

1. **Public Health England.** 2019. English Surveillance Programme for Antimicrobial Utilisation and Resistance (ESPAUR) Report 2018–2019. Public Health England.

2. **Tenaillon O, Skurnik D, Picard B, Denamur E.** 2010 Mar. The population genetics of commensal Escherichia coli. Nat. Rev. Microbiol. 8(3):207–217. http://www.nature.com/articles/nrmicro2298.

3. **The Global Sewage Surveillance project consortium, Hendriksen RS, Munk P, Njage P, van Bunnik B, McNally L, Lukjancenko O, Röder T, Nieuwenhuijse D, Pedersen SK, Kjeldgaard J, Kaas RS, Clausen PTLC, Vogt JK, Leekitcharoenphon P, van de Schans MGM, Zuidema T, de Roda Husman AM, Rasmussen S, Petersen B, Amid C, Cochrane G, Sicheritz-Ponten T, Schmitt H, Alvarez JRM, Aidara-Kane A, Pamp SJ, Lund O, Hald T, Woolhouse M, Koopmans MP, Vigre H, Petersen TN, Aarestrup FM.** 2019 Dec. Global monitoring of antimicrobial resistance based on metagenomics analyses of urban sewage. Nat. Commun. 10(1):1124. http://www.nature.com/articles/s41467-019-08853-3.

4. **Public Health England.** 2017. Preventing healthcare associated Gram-negative bloodstream infections: an improvement resource. Public Health England.

5. **Kizny Gordon AE, Mathers AJ, Cheong EYL, Gottlieb T, Kotay S, Walker AS, Peto TEA, Crook DW, Stoesser N.** 2017 May. The Hospital Water Environment as a Reservoir for Carbapenem-Resistant Organisms Causing Hospital-Acquired Infections—A Systematic Review of the Literature. Clin. Infect. Dis. 64(10):1435–1444. https://academic.oup.com/cid/article-lookup/doi/10.1093/cid/cix132.

6. **Mathers AJ, Crook D, Vaughan A, Barry KE, Vegesana K, Stoesser N, Parikh HI, Sebra R, Kotay S, Walker AS, Sheppard AE.** 2019 Mar. Klebsiella quasipneumoniae Provides a Window into Carbapenemase Gene Transfer, Plasmid Rearrangements, and Patient Interactions with the Hospital Environment. Antimicrob. Agents Chemother. 63(6):e02513–18, /aac/63/6/AAC.02513–18.atom. http://aac.asm.org/lookup/doi/10.1128/AAC.02513-18.

7. **Hopman J, Tostmann A, Wertheim H, Bos M, Kolwijck E, Akkermans R, Sturm P, Voss A, Pickkers P, vd Hoeven H.** 2017 Dec. Reduced rate of intensive care unit acquired gram-negative bacilli after removal of sinks and introduction of 'water-free' patient care. Antimicrob. Resist. & Infect. Control. 6(1):59. http://aricjournal.biomedcentral.com/articles/10.1186/s13756-017-0213-0.

8. **Mathers AJ, Vegesana K, German Mesner I, Barry KE, Pannone A, Baumann J, Crook DW, Stoesser N, Kotay S, Carroll J, Sifri CD.** 2018 Jul. Intensive Care Unit Wastewater Interventions to Prevent Transmission of Multispecies Klebsiella pneumoniae Carbapenemase–Producing Organisms. Clin. Infect. Dis. 67(2):171–178. https://academic.oup.com/cid/article/67/2/171/4834952.

9. **Quainoo S, Coolen JPM, van Hijum SAFT, Huynen MA, Melchers WJG, van Schaik W, Wertheim HFL.** 2017 Oct. Whole-Genome Sequencing of Bacterial Pathogens: the Future of Nosocomial Outbreak Analysis. Clin. Microbiol. Rev. 30(4):1015–1063. http://cmr.asm.org/lookup/doi/10.1128/CMR.00016-17.

10. **Quince C, Walker AW, Simpson JT, Loman NJ, Segata N.** 2017 Sep. Shotgun metagenomics, from sampling to analysis. Nat. Biotechnol. 35(9):833–844. http://www.nature.com/articles/nbt.3935.

11. **Sukhum KV, Diorio-Toth L, Dantas G.** 2019 Sep. Genomic and Metagenomic Approaches for Predictive Surveillance of Emerging Pathogens and Antibiotic Resistance. Clin. Pharmacol. & Ther. 106(3):512–524. https://onlinelibrary.wiley.com/doi/abs/10.1002/cpt.1535.

12. **on behalf of the REHAB consortium, Gweon HS, Shaw LP, Swann J, De Maio N, AbuOun M, Niehus R, Hubbard ATM, Bowes MJ, Bailey MJ, Peto TEA, Hoosdally SJ, Walker AS, Sebra RP, Crook DW, Anjum MF, Read DS, Stoesser N.** 2019 Dec. The impact of sequencing depth on the inferred taxonomic composition and AMR gene content of metagenomic samples. Environ. Microbiome 14(1):7. https://environmentalmicrobiome.biomedcentral.com/articles/10.1186/s40793-019-0347-1.

13. **Johnson RC, Deming C, Conlan S, Zellmer CJ, Michelin AV, Lee-Lin S, Thomas PJ, Park M, Weingarten RA, Less J, Dekker JP, Frank KM, Musser KA, McQuiston JR, Henderson DK, Lau AF, Palmore TN, Segre JA.** 2018 Dec. Investigation of a Cluster of Sphingomonas koreensis Infections. New Engl. J. Medicine 379(26):2529–2539. http://www.nejm.org/doi/10.1056/NEJMoa1803238.

14. **Carattoli A, Villa L, Feudi C, Curcio L, Orsini S, Luppi A, Pezzotti G, Magistrali CF.** 2017 Aug. Novel plasmid-mediated colistin resistance mcr-4 g ene in Salmonella and Escherichia coli , Italy 2013, Spain and Belgium, 2015 to 2016. Eurosurveillance 22(31):30589. http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=22851.

15. **Bush SJ, Foster D, Eyre DW, Clark EL, De Maio N, Shaw LP, Stoesser N, Peto TEA, Crook DW, Walker AS.** 2019. Genomic diversity affects the accuracy of bacterial SNP calling pipelines. bioRxiv https://www.biorxiv.org/content/early/2019/05/31/653774.

16. **Zhi S, Banting G, Stothard P, Ashbolt NJ, Checkley S, Meyer K, Otto S, Neumann NF.** 2019 Jun. Evidence for the evolution, clonal expansion and global dissemination of water treatment-resistant naturalized strains of Escherichia coli in wastewater. Water Res. 156:208–222. https://linkinghub.elsevier.com/retrieve/pii/S004313541930243X.

17. **Lowe C, Willey B, O'Shaughnessy A, Lee W, Lum M, Pike K, Larocque C, Dedier H, Dales L, Moore C, McGeer A, the Mount Sinai Hospital Infection Control Team.** 2012 Aug. Outbreak of Extended-Spectrum β-Lactamase–producing Klebsiella oxytoca Infections Associated with Contaminated Handwashing Sinks1. Emerg. Infect. Dis. 18(8):1242–1247. http://wwwnc.cdc.gov/eid/article/18/8/11-1268_article.htm.

18. **Kotay SM, Donlan RM, Ganim C, Barry K, Christensen BE, Mathers AJ.** 2018 Oct. Droplet- Rather than Aerosol-Mediated Dispersion Is the Primary Mechanism of Bacterial Transmission from Contaminated Hand-Washing Sink Traps. Appl. Environ. Microbiol. 85(2):e01997–18, /aem/85/2/AEM.01997–18.atom. http://aem.asm.org/lookup/doi/10.1128/AEM.01997-18.

19. **Aranega-Bou P, George RP, Verlander NQ, Paton S, Bennett A, Moore G, Aiken Z, Akinremi O, Ali A, Cawthorne J, Cleary P, Crook DW, Decraene V, Dodgson A, Doumith M, Ellington M, Eyre DW, George RP, Grimshaw J, Guiver M, Hill R, Hopkins K, Jones R, Lenney C, Mathers AJ, McEwan A, Moore G, Neilson M, Neilson S, Peto TEA, Phan HTT, Regan M, Seale AC, Stoesser N, Turner-Gardner J, Watts V, Walker J, Sarah Walker A, Wyllie D, Welfare W, Woodford N.** 2019 May. Carbapenem-resistant Enterobacteriaceae dispersal from sinks is linked to drain position and drainage rates in a laboratory model system. J. Hosp. Infect. 102(1):63–69. https://linkinghub.elsevier.com/retrieve/pii/S0195670118307114.

20. **Carling PC.** 2018 Aug. Wastewater drains: epidemiology and interventions in 23 carbapenem-resistant organism outbreaks. Infect. Control. & Hosp. Epidemiol. 39(8):972–979. https://www.cambridge.org/core/product/identifier/S0899823X18001381/type/journal_article.

21. **Chau K.**, WGS investigation of three OXA-48 carbapenemase bacteraemias and sink colonisation on a haematology unit. Madrid, Spain 2018.

22. **Public Health England.** 2014. Standards for microbiology investigations (UK SMI). Public Health England.

23. **Wood DE, Salzberg SL.** 2014. Kraken: ultrafast metagenomic

Multi-omic surveillance of *Escherichia coli* and *Klebsiella* spp. in hospital sink drains and patients

sequence classification using exact alignments. Genome Biol. 15(3):R46. http://genomebiology.biomedcentral.com/articles/10.1186/gb-2014-15-3-r46.

24. **Seemann T.**, Snippy 2019. https://github.com/tseemann/snippy, original-date: 2014-05-15T01:56:12Z.

25. **Seemann T.**, Shovill 2019. https://github.com/tseemann/shovill, original-date: 2016-09-05T23:50:20Z.

26. **Eyre D.**, runListCompare 2019. https://github.com/davideyre/runListCompare, original-date: 2017-12-15T08:49:46Z.

27. **Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ.** 2015 Jan. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. Mol. Biol. Evol. 32(1):268–274. https://academic.oup.com/mbe/article-lookup/doi/10.1093/molbev/msu300.

28. **Didelot X, Wilson DJ.** 2015 Feb. ClonalFrameML: Efficient Inference of Recombination in Whole Bacterial Genomes. PLOS Comput. Biol. 11(2):e1004041. https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1004041.

29. **Ondov BD, Starrett GJ, Sappington A, Kostic A, Koren S, Buck CB, Phillippy AM.** 2019 Nov. Mash Screen: high-throughput sequence containment estimation for genome discovery. Genome Biol. 20(1):232. https://doi.org/10.1186/s13059-019-1841-x.

30. **Simonsen M, Mailund T, Pedersen CNS.** 2008. Rapid Neighbour-Joining. In: Crandall KA, Lagergren J, editors. Algorithms in Bioinformatics, vol. 5251 Berlin, Heidelberg: Springer Berlin Heidelberg, p. 113–122. http://link.springer.com/10.1007/978-3-540-87361-7_10.

31. **Lees JA, Harris SR, Tonkin-Hill G, Gladstone RA, Lo SW, Weiser JN, Corander J, Bentley SD, Croucher NJ.** 2019 Feb. Fast and flexible bacterial genomic epidemiology with PopPUNK. Genome Res. 29(2):304–316. http://genome.cshlp.org/lookup/doi/10.1101/gr.241455.118.

32. **Hunt M, Mather AE, Sánchez-Busó L, Page AJ, Parkhill J, Keane JA, Harris SR.** 2017 Oct. ARIBA: rapid antimicrobial resistance genotyping directly from sequencing reads. Microb. Genomics 3(10). https://www.microbiologyresearch.org/content/journal/mgen/10.1099/mgen.0.000131.

33. **Alcock BP, Raphenya AR, Lau TTY, Tsang KK, Bouchard M, Edalatmand A, Huynh W, Nguyen ALV, Cheng AA, Liu S, Min SY, Miroshnichenko A, Tran HK, Werfalli RE, Nasir JA, Oloni M, Speicher DJ, Florescu A, Singh B, Faltyn M, Hernandez-Koutoucheva A, Sharma AN, Bordeleau E, Pawlowski AC, Zubyk HL, Dooley D, Griffiths E, Maguire F, Winsor GL, Beiko RG, Brinkman FSL, Hsiao WWL, Domselaar GV, McArthur AG.** 2019 Oct. CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database. Nucleic Acids Res. p. gkz935. https://academic.oup.com/nar/advance-article/doi/10.1093/nar/gkz935/5608993.

34. **Galili T.** 2015 Nov. dendextend: an R package for visualizing, adjusting and comparing trees of hierarchical clustering. Bioinformatics 31(22):3718–3720. https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/btv428.

35. **Wood DE, Lu J, Langmead B.** 2019 Dec. Improved metagenomic analysis with Kraken 2. Genome Biol. 20(1):257. https://genomebiology.biomedcentral.com/articles/10.1186/s13059-019-1891-0.

36. **Li D, Liu CM, Luo R, Sadakane K, Lam TW.** 2015 May. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics 31(10):1674–1676. https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/btv033.

37. **Li H.** 2018 Sep. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 34(18):3094–3100. https://academic.oup.com/bioinformatics/article/34/18/3094/4994778.

38. **Constantinides B, L Robertson D.** 2017 Jul. Kindel: indel-aware consensus for nucleotide sequence alignments. The J. Open Source Softw. 2(15):282. http://joss.theoj.org/papers/10.21105/joss.00282.

39. **Breitwieser FP, Salzberg SL.** 2019 Sep. Pavian: interactive analysis of metagenomics data for microbiome studies and pathogen identification. Bioinformatics p. btz715. https://academic.oup.com/bioinformatics/advance-article/doi/10.1093/bioinformatics/btz715/5573755.

40. **SciPy 1 0 Contributors, Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, Burovski E, Peterson P, Weckesser W, Bright J, van der Walt SJ, Brett M, Wilson J, Millman KJ, Mayorov N, Nelson ARJ, Jones E, Kern R, Larson E, Carey CJ, Polat Feng Y, Moore EW, VanderPlas J, Laxalde D, Perktold J, Cimrman R, Henriksen I, Quintero EA, Harris CR, Archibald AM, Ribeiro AH, Pedregosa F, van Mulbregt P.** 2020 Feb. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat. Methods http://www.nature.com/articles/s41592-019-0686-2.

41. **Perez F, Granger BE.** 2007. IPython: A System for Interactive Scientific Computing. Comput. Sci. & Eng. 9(3):21–29. http://ieeexplore.ieee.org/document/4160251/.

42. **Hunter JD.** 2007. Matplotlib: A 2D Graphics Environment. Comput. Sci. & Eng. 9(3):90–95. http://ieeexplore.ieee.org/document/4160265/.

43. **Argimón S, Abudahab K, Goater RJE, Fedosejev A, Bhai J, Glasner C, Feil EJ, Holden MTG, Yeats CA, Grundmann H, Spratt BG, Aanensen DM.** 2016 Nov. Microreact: visualizing and sharing data for genomic epidemiology and phylogeography. Microb. Genomics 2(11). https://www.microbiologyresearch.org/content/journal/mgen/10.1099/mgen.0.000093.