Epidemiology of paediatric gastrointestinal colonisation by extended spectrum cephalosporin-resistant *Escherichia coli* and *Klebsiella pneumoniae* isolates in north-west Cambodia

JJ van Aartsen[1,2], CE Moore[1], CM Parry[3], P Turner[4,5], N Phot[6], S Mao[6], K Suy[6], T Davies[1], A Giess[1], AE Sheppard[1], TEA Peto[1], NPJ Day[5,7], DW Crook[1], AS Walker[1], N Stoesser[1#]

[1] Nuffield Department of Clinical Medicine and the National Institute for Health Research Oxford Biomedical Research Centre (NIHR-OxBRC), University of Oxford, Oxford, United Kingdom.

[2] Department of Clinical Infection, Microbiology and Immunology, Institute of Infection and Global Health, University of Liverpool, Liverpool, United Kingdom.

[3] Clinical Sciences, Liverpool School of Tropical Medicine, Liverpool, UK

[4] Cambodia-Oxford Medical Research Unit, Angkor Hospital for Children, Siem Reap, Cambodia.

[5] Centre for Tropical Medicine and Global Health, Nuffield Department of Medicine, University of Oxford, Oxford, UK

[6] Angkor Hospital for Children, Siem Reap, Cambodia.

[7] Mahidol-Oxford Tropical Medicine Research Unit, Faculty of Tropical Medicine, Mahidol University, Bangkok, Thailand.

Running title: Paediatric ESBL Enterobacteriaceae carriage, Cambodia

24

Corresponding author: Department of Microbiology/Infectious Diseases, John Radcliffe Hospital, Headley Way, Headington, OX3 9DU, United Kingdom. Tel: +44-(0)1865 220856; E-mail: nicole.stoesser@ndm.ox.ac.uk

ABSTRACT

Extended-spectrum cephalosporin resistance (ESC-R) in *Escherichia coli* and *Klebsiella pneumoniae* is a healthcare threat; high gastrointestinal carriage rates are reported from South-east Asia. Colonisation prevalence data in Cambodia are lacking. We determined gastrointestinal colonisation prevalence of ESC-resistant *E. coli* (ESC-R-EC) and *K. pneumoniae* (ESC-R-KP) in Cambodian children/adolescents and associated risk factors; characterised relevant resistance genes, their genetic contexts, and the genetic relatedness of ESC-R strains using whole genome sequencing (WGS). Faeces and questionnaire data were obtained from individuals <16 years in northwestern Cambodia, 2012. WGS of cultured ESC-R-EC/KP was performed (Illumina). Maximum likelihood phylogenies were used to characterise relatedness of isolates; ESC-R-associated resistance genes and their genetic contexts were identified from *de novo* assemblies using BLASTn and automated/manual annotation. 82/148 (55%) of children/adolescents were ESC-R-EC/KP colonised; 12/148 (8%) were co-colonised with both species. Independent risk factors for colonisation were hospitalisation (OR: 3.12, 95%, CI [1.52-6.38]) and intestinal parasites (OR: 3.11 [1.29-7.51]); school attendance conferred decreased risk (OR: 0.44 [0.21-0.92]. ESC-R strains were diverse; the commonest ESC-R mechanisms were $bla_{\text{CTX-M}}$ 1 and 9 sub-family variants. Structures flanking these genes were highly variable, and for $bla_{\text{CTX-M-15, -55 and -27}}$, frequently involved IS*26*. Chromosomal $bla_{\text{CTX-M}}$ integration was common in *E. coli*. Gastrointestinal ESC-R-EC/KP colonisation is widespread in Cambodian children/adolescents; hospital admission and intestinal parasites are independent risk factors. The genetic contexts of $bla_{\text{CTX-M}}$ are highly mosaic, consistent with rapid horizontal exchange. Chromosomal integration of $bla_{\text{CTX-M}}$ may result in stable propagation in these community-associated pathogens.

52 **MAIN TEXT**

53 **INTRODUCTION**

54 *Escherichia coli* and *Klebsiella pneumoniae* are two bacterial pathogens of the
55 Enterobacteriaceae family that can cause a wide spectrum of clinical disease, ranging from
56 cystitis and intra-abdominal abscesses to sepsis. Both species also asymptomatically colonise
57 the gastrointestinal tract, a reservoir that assists in the acquisition and spread of antimicrobial
58 resistance (AMR)(1, 2). The increasing prevalence of AMR worldwide is reducing the
59 efficacy of our limited armamentarium of empirical broad-spectrum antibiotics, such as
60 extended-spectrum cephalosporins (ESCs), resulting in increased healthcare costs and
61 mortality(3-5).

62

63 Recent reports from South-east Asia show substantial variation between country and cohort
64 in gastrointestinal colonisation by Enterobacteriaceae possessing Ambler class A extended
65 spectrum beta-lactamases (ESBLs) and/or class C AmpC enzymes, which can hydrolyse third
66 and fourth generation cephalosporins. In the Lao People's Democratic Republic, for example,
67 23% of pre-school children carried these strains, in contrast to a much higher prevalence of
68 65.7% in a rural Thai adult population(6-8). Data describing the prevalence and mechanisms
69 of antibiotic resistance in Cambodia are limited to only a few studies. Vlieghe and colleagues
70 found 49.7% of Enterobacteriaceae from blood cultures in Phnom Penh from 2007-2010 were
71 cefotaxime-resistant, mostly due to CTX-M-15 and CTX-M-14 enzymes(9). Studies from
72 2004/5 and 2007-2011 identified ESC resistance in 36-44% of urinary tract infection
73 isolates(10, 11). The gastrointestinal colonisation prevalence of ESC-resistant (ESC-R) *E.*
74 *coli* and *K. pneumoniae* in Cambodia has previously only been investigated in hospitalised
75 neonates(12).

This study aimed to: (i) estimate the prevalence of gastrointestinal colonisation with ESC-resistant *E. coli* (ESC-R-EC) and *K. pneumoniae* (ESC-R-KP) in Cambodian children and adolescents, and the molecular mechanisms responsible; (ii) investigate risk factors for ESC-R colonisation; (iii) determine genetic relatedness of ESC-R strains.

RESULTS

Sampling, culture and basic demographics

In total, 196 faecal samples were obtained from a consecutive subset of children/adolescents enrolled in a helminth prevalence study. 48 samples were excluded from this study because of: (i) lack of specific consent for wider use of the faecal samples beyond the helminth survey (n=36); (ii) no epidemiological data records (n=1); (iii) no (n=3) or poor (n=5) growth on culture; or (iv) replicate samples for the same patient (n=3), leaving 148 samples/individuals for analysis.

Overall, 184 distinct colony types grew within the cefpodoxime inhibition zones; 141 were pink (presumed *E. coli*) and 43 were blue (presumed *Klebsiella* spp., *Enterobacter* spp. or *Citrobacter* spp.). All pink colonies but only 22/43 (54%) blue colonies were confirmed as phenotypically ESC-R using BSAC methods. All 163 confirmed ESC-R isolates were sequenced; two failed and were excluded from further analysis. Of the 161 sequences, *in silico* species identification confirmed 135 (84%) isolates were *E. coli*, 18 (11%) *K. pneumoniae*, and 8 (5%) *Enterobacter* spp. 38 *E. coli* isolates and one *K. pneumoniae* isolate were genetically sufficiently closely related to another isolate obtained from the same patient sample to be considered as the same strain (defined as ≤5 chromosomal SNVs); these were also excluded leaving 122 isolates for analysis. None of the 148 faecal samples yielded imipenem resistant colonies.

Participants were median 4.2 years old (interquartile range: 1.1-8.8) at sample collection; 70/148 (47%) were male. 70/147 (48%; 1 missing) were inpatients at sample collection. Although most were from Siem Reap province (99/148 [67%]), the hospital catchment is such that the remainder were recruited from 10 other provinces. 16/148 (11%) were clinically malnourished, and 23/148 (16%) had ≥1 underlying chronic medical condition including HIV (n=5), haematological disease (n=3), congenital cardiac disease (n=5), tuberculosis (n=4), and asthma (n=2)(Table 1).

Prevalence of and risk factors for colonisation with ESC-R EC and/or ESC-R-KP

A total of 114 confirmed ESC-R-EC (n=97) and ESC-R-KP (n=17) remained in the analysis and were carried by 82/148 participants, giving a combined ESC-R-EC/KP prevalence of 55% (95% CI: 47%-64%); 53% for ESC-R EC (79/148 patients; 95% CI: 45%-62%) and 10% for ESC-R KP (15/148 patients; 95% CI: 6%-16%). Co-colonisation with both ESC-R-EC and ESC-R-KP was observed in 12/82 (15%). Independent risk factors for ESC-R-EC/KP colonisation included being a current inpatient (OR=3.64; 95% CI [1.71-7.74), p=0.001) and the presence of faecal parasites (OR=3.96 [1.55-10.13], p=0.004). ESC-R-EC/KP colonisation was lower in males (OR=0.39 [0.18-0.84], p=0.015) and in those attending school (OR=0.39 [0.18-0.83], p=0.015)(Table 1).

Sequence type, Ambler class and genetic mechanisms of ESC-R

The 97 ESC-R-EC isolates came from 33 known and 6 novel STs (Fig.1, for details see Table S1). 22% (17/79) of patients were colonised by at least two different ESC-R-EC STs, although this may underestimate diversity as only a small number of colonies (≤3) were

sampled per patient(25). The 17 ESC-R-KP strains came from 11 known and 3 novel STs (n=4 isolates) (Fig.2, Table S2). Two patients were colonised by two different ESC-R *K. pneumoniae* STs (2/15, 13%).

In total, 77% (88/114) and 23% (26/114) of isolates displayed Ambler class A or C phenotypes, respectively. Neither species were associated with Ambler class A (76% [74/97] versus 82% [14/17]) or class C (23% [23/97] versus 18% [3/17]; Fishers exact test; p=0.759). In all class A isolates the phenotype could be explained by the presence of one (84/88, 95%) or two (4/88, 5%) $bla_{\text{CTX-M}}$ genes; bla_{SHV} (12/88, 12%) and bla_{VEB} (1/88, 1%) occurred less commonly. Class C gene families were only identified in 39% (10/26) of phenotypically class C isolates: specifically $bla_{\text{CMY-2}}$ (8/26, 31%) or bla_{DHA} (2/26, 8%). In the remaining 16 isolates, the genetic basis for the class C phenotype was unclear; of note, however, *ampC* promoter mutations were not assessed. 111 $bla_{\text{CTX-M}}$ genes were found in 94% (107/114) of ESC-R-EC/KP, with two separate alleles identified in 4% of isolates (4/114). The most frequently identified allele was $bla_{\text{CTX-M-15}}$ (53/111, 48%), followed by: $bla_{\text{CTX-M-55}}$ (24/111, 22%), $\text{bla}_{\text{CTX-M-14}}$ (17/111,15%), $bla_{\text{CTX-M-27}}$ (14/111, 13%) and $bla_{\text{CTX-M-24}}$ (3/111, 3%). Two different $bla_{\text{CTX-M}}$ alleles were found in 21% (18/82) of individuals carrying ESC-R-EC/KP. bla_{SHV} genes were identified in 15/17 *K. pneumoniae*, including $bla_{\text{SHV-1/ SHV-1-like}}$ (3/15, 20%), $\text{bla}_{\text{SHV-11/ SHV-11-like}}$ (4/15, 27%), $bla_{\text{SHV-27-like}}$ (1/15, 7%), $bla_{\text{SHV-28}}$ (1/15, 7%), $bla_{\text{SHV-33}}$ (3/15, 20.0%) and $bla_{\text{SHV-83}}$ (1/15, 7%), $bla_{\text{SHV-99-like}}$ (1/15, 7%) and $bla_{\text{SHV-142}}$ (1/15, 7%). All eight $bla_{\text{CMY-2/CMY-2-like}}$ genes were found in *E. coli*. The study population carriage prevalence of common genetic mechanisms encoded by ESC-R EC/KP was therefore: 53% $bla_{\text{CTX-M}}$ (78/148), 9% bla_{SHV} (14/148), 1% bla_{VEB} (1/148), 5% $bla_{\text{CMY-2}}$ (8/148), 1% bla_{DHA} (2/148). Two individuals (1%) carried isolates with $bla_{\text{OXA-48}}$ (one *K. pneumoniae* ST48 and one *E. coli* ST648); no other carbapenem resistance mechanisms were identified.

Genetic context of ESC-R genes

For the 41 *E. coli* harbouring *bla*$_{\text{CTX-M-15}}$, it was chromosomally located in five cases (12%), and likely in plasmid contexts in two; in the remaining cases it was not possible to determine wider chromosomal/plasmid location (Table 2). One isolate (38P1) harboured short contigs containing truncated *bla*$_{\text{CTX-M-15}}$, leaving 40 cases in which to evaluate the immediate flanking contexts surrounding the *bla*$_{\text{CTX-M}}$ gene. All contained IS*Ecp1* upstream of *bla*$_{\text{CTX-M-15}}$, but with considerable evidence of additional mobilisation events/mosaicism (Table 2). In particular, IS*Ecp1* was truncated by IS*26* at 24, 497, 524, 1067, 1173, 1421, or 1489bp in 13 isolates, consistent with at least seven IS*26*-associated insertion events within IS*Ecp1*(Fig.3). Another 13 IS*Ecp1* elements were truncated by contig breaks, without any specific associated genetic signatures, although contig breaks are frequently due to repeat structures and may therefore have represented additional disruption events. One isolate had an intact IS*Ecp1* element, without any wider flanking upstream context. The 13 cases with an intact IS*Ecp1* were consistently flanked by variable lengths of Tn*2*, which was truncated by an IS*26* right IRR in 2/7 evaluable cases (and by an unknown sequence in the other 5/7). Two isolates had a complete Tn*2* structure interrupted by IS*Ecp1*-*bla*$_{\text{CTX-M-15}}$ (TCTCA-TCTCA and TTTTA-TAAAA target site sequences [TSSs] respectively)(Fig.3). Overall, genetic contexts of *bla*$_{\text{CTX-M-15}}$ were consistent with integration and mobilisation of IS*Ecp1*-*bla*$_{\text{CTX-M-15}}$ within a Tn*2* element, as previously described(26), with subsequent rearrangement events facilitated by IS*26* and perhaps other ISs(27)(Table S3).

For the 24 *E. coli* harbouring *bla*$_{\text{CTX-M-55}}$, it was chromosomally located in 4 (17%), plasmid in 3 (13%) and unknown in 16 (67%). One contig contained a truncated *bla*$_{\text{CTX-M-55}}$, leaving 23 evaluable contexts. Similar to *bla*$_{\text{CTX-M-15}}$, it was invariably associated with IS*Ecp1*

upstream of *bla*$_{\text{CTX-M-55}}$ (Fig.4), which was often incomplete, representing at least 3 different IS*26*-associated IS*Ecp1* disruption events (Table 2). Intact IS*Ecp1* were flanked by variable lengths of Tn*2* sequence, apart from 120P1 where the contig was truncated immediately at the 5' end of IS*Ecp1*. One isolate (2P1) had the same *bla*$_{\text{CTX-M}}$/Tn*2* unit as for *bla*$_{\text{CTX-M-15}}$ (but with TACTC-TAAAA), consistent with the evolution of *bla*$_{\text{CTX-M-55}}$ from *bla*$_{\text{CTX-M-15}}$ (1 SNV difference) within this unit (Figs.3, 4).

For the 15 *E. coli* harbouring *bla*$_{\text{CTX-M-14}}$, it was chromosomally located in 2 (13%) cases, plasmid-associated in 5 (33%), and unknown in 8 (53%). Again, it was invariably associated with IS*Ecp1*, but more often complete and with different mechanisms of disruption (2 IS*Vsa5-like* sequence, one IS*1S* R IRR). All cases had an IS*903* element at the 3' end of *bla*$_{\text{CTX-M-14}}$; this had been disrupted in 6 cases, with additional contig breaks in 5 cases (Fig.5). Two of three *E. coli bla*$_{\text{CTX-M-24}}$ contexts were chromosomal, with flanking contexts similar to *bla*$_{\text{CTX-M-14}}$ (Fig.S1). In the 12 *bla*$_{\text{CTX-M-27}}$ cases, the *ISEcp1* element had been disrupted by an IS*26* L IRR in all contexts, at 149, 192, 208 and 388bp, but the wider genetic context of this structure was indeterminable in all cases (Fig.S2).

Overall, *bla*$_{\text{CTX-M}}$ was chromosomal in 13/92 cases (14%; 13/25 [52%] cases where plasmid versus chromosomal location could be assessed), suggesting that CTX-M genes may be incorporated chromosomally and indiscriminately in significant numbers of colonising *E. coli*, with possible implications for their stable propagation within the wider *E. coli* population.

198 For *K. pneumoniae*, 12 isolates harboured *bla*$_{\text{CTX-M-15}}$, in a plasmid-associated context in 9/12
199 cases, and an unknown context in 3/12 cases. Three isolates harboured a complete *bla*$_{\text{CTX-M-15}}$/Tn*2* complex with GTTAA-GTTAA TSS, most consistent with a direct transposition of
200
201 this element into a plasmid context. In the other isolates, the IS*Ecp1*-*bla*$_{\text{CTX-M-15}}$-ORF477 was
202 flanked by variable stretches of Tn*2*-associated sequence identical to that found in the *E. coli*
203 isolates, and similarly truncated either as a result of contig breaks, or by IS*26* inverted
204 repeats, consistent with between species and within species mobilisation (Fig.S3).

205

206 Four *K. pneumoniae* isolates harboured *bla*$_{\text{CTX-M-9}}$ group genes; two of these (*bla*$_{\text{CTX-M-14}}$)
207 shared the same IS*Ecp1* (Fig.5) and ~18kb upstream flanking plasmid sequence; and two
208 (*bla*$_{\text{CTX-M-27}}$) an IS*Ecp1* element truncated at position 1499 by an IS*26* L IRR (Fig.S2).

209

210 **DISCUSSION**

211 We observed significant gastrointestinal carriage prevalence of both ESC-R-EC and ESC-R-
212 KP in Cambodian children sampled in 2012; approximately one in twelve children was co-
213 colonised with ESC-R strains of both species. A wide diversity of ESC-R strain types was
214 observed, including several genotypes categorised as “high risk” clones, such as *E. coli* STs
215 38, 405, 131, 354 and 648(13). The predominant ESC-R genotypic mechanism was *bla*$_{\text{CTX-M}}$,
216 with the major allelic variants being those widely described elsewhere in Asia (Group 1:
217 *bla*$_{\text{CTX-M-15, -55}}$, Group 9: *bla*$_{\text{CTX-M-14, -24, -27}}$). Approximately one-third of the Cambodian
218 population is <18 years old, so this group may be acting as a significant reservoir for the
219 spread of anti-microbial resistant organisms. We did not observe particularly high rates of
220 colonisation with carbapenem-resistant isolates (2 (1%) individuals), but one of these was an
221 out-patient, with an OXA-48 *E. coli* isolate, and without any known chronic health problems,

222 suggesting that there may be some carriage of carbapenem-resistant isolates in the
223 community. Further assessment of the extent of carriage of carbapenem-resistant EC/KP in
224 this context is warranted.

225

226 Independent risk factors for colonisation included inpatient status, consistent with
227 transmission within hospital, and/or selection of these organisms from low-level carriage by
228 the use of antibiotics on admission given the high burden of infectious diseases in this region.
229 Infection control (IC) in resource-limited settings remains challenging, and despite
230 improvements within the study hospital(14), recent longitudinal surveillance within the
231 neonatal care unit identified high rates of import of ESC-R-KP (62% colonised on admission)
232 as well as nosocomial acquisition (23%)(12). In-patient acquisition of ESC-R-EC/KP has also
233 been identified as a major problem in other low/middle-income settings(15). The specific
234 effect of faecal parasites on gut microbiota is not well-studied, but they are thought to
235 significantly perturb microbial diversity(16). Helminth infestation may also result in
236 inappropriate antimicrobial use, including antibiotics, perhaps leading to secondary
237 colonization with drug-resistant commensals. The decreased risk associated with school
238 attendance has been observed in a previous study in Spain(17), and may represent a proxy
239 marker for increased socio-economic status, and parental levels of education, which were not
240 evaluated here, but may translate into better awareness of appropriate antibiotic use(18, 19).
241 The decreased risk associated with male gender is unexplained; but independent associations
242 for ESBL-EC/KP colonization have been described for both genders in previous studies(15,
243 20, 21).

244

Of particular importance was the high prevalence of chromosomal integration of $bla_{\text{CTX-M}}$ in *E. coli* in this study (>14%), perhaps contributing to the stable propagation of this resistance gene family within certain strains. Chromosomal integration of $bla_{\text{CTX-M}}$ in *K. pneumoniae* was not observed in our study, although it has been seen in Spain(22). In addition, despite the limitations of short-read assemblies, the genetic contexts of $bla_{\text{CTX-M}}$ suggested high levels of genetic plasticity in flanking structures, and significant associations with IS*26* for $bla_{\text{CTX-M-15}}$, $bla_{\text{CTX-M-55}}$, and $bla_{\text{CTX-M-27}}$. IS*26* has been previously hypothesised to facilitate the mobility of $bla_{\text{CTX-M}}$ and genetic rearrangement of resistance gene plasmids, and is likely contributing to the dissemination of these resistance genes within the human gastrointestinal reservoir(23-25).

This study has several limitations. Our survey dates from 2012, and the epidemiology of ESC-R EC/KP carriage may have changed in the intervening timeframe; nevertheless, our data represent the largest molecular epidemiological study of gastrointestinal ESC-R-EC/KP colonisation in Cambodia and a useful benchmark for future studies. We only included up to three bacterial colonies per faecal sample, likely resulting in significant under-estimation of the diversity present at the population level(26). Short-read sequencing resulted in limited information regarding the wider genetic context of important resistance genes conferring ESC-R; nevertheless, we were still able to ascertain that the genetic contexts of these resistance genes are extremely diverse. Our outpatient study population may not be truly representative of healthy children in the community, given that these individuals had presented to the outpatient department for some form of medical review. Lack of more detailed information on some potential risk factors meant we were unable to fully assess the specific mechanisms promoting ESC-R EC/KP colonisation. Further work characterising the role of healthcare admissions, socio-economic factors and intestinal parasites on the

acquisition and long-term carriage dynamics of these strains would be valuable. In addition, our sample size was too small and sparse to investigate geographical clustering of strain types, and to investigate specific risk factors for colonisation with common strain types or resistance gene alleles.

Despite these limitations, our study adds to the growing body of literature demonstrating widespread gastrointestinal colonisation with ESC-R-EC and ESC-R KP in Southeast Asia(8), and showing that exposure to this reservoir may in turn act as a source for the wider, global transfer of these strains(27). The genetic contexts of important resistance genes are highly mosaic, consistent with rapid exchange of resistance genes within and between bacterial hosts. Significant levels of chromosomal integration of the most important ESC-R gene family, $bla_{\text{CTX-M}}$, were also observed, and may result in these genes being stably maintained and propagated in one of the most common community-associated pathogens, namely *E. coli*.

MATERIALS AND METHODS

Patients and setting

Faecal samples were obtained from a consecutive subset of children/adolescents (<16 years) who had been enrolled prospectively in a helminth prevalence study at Angkor Hospital for Children in Siem Reap, Cambodia, from 3rd April 2012 to 29th June 2012, as described in(28).

Microbiological methods

Samples were frozen at -80°C as aliquots homogenised in 0.9% sterile saline with 10% glycerol within an hour of receipt in the laboratory. For this study, faecal samples were thawed, and aliquots diluted 1:10 in saline and incubated for 16 hours at 37°C on Orientation CHROMagar (BD, Oxford, United Kingdom) with 10 µg cefpodoxime and 10 µg imipenem discs (Oxoid, Basingstoke, United Kingdom). For each faecal sample, up to three pink and/or dark blue colonies with different colonial morphotypes that grew within the cefpodoxime zone of inhibition (presumed ESC-R-EC and ESC-R-KP respectively) were selected for further analysis. Each selected colony was tested using the British Society of Antimicrobial Chemotherapy (BSAC) combination disc method to identify whether cefpodoxime (ESC) resistance was mediated via ESBLs (Class A: cefpodoxime-resistant, and cefpodoxime+clavulanic acid-sensitive) or via non-ESBL mechanisms (e.g. Class C AmpC beta-lactamases: cefpodoxime-resistant, and cefpodoxime+clavulanic acid-resistant)(29). All identified ESC-R colonies were stored frozen at -80°C in nutrient broth with 10% glycerol.

Whole genome sequencing and sequence data processing

DNA was extracted from sub-cultured ESC-R isolates using a commercial kit (Fujifilm Quickgene, Japan) with an additional mechanical lysis step (Fastprep MP Biomedicals, USA). All isolates were sequenced using the Illumina HiSeq 2500, generating 150bp paired-end reads. Sequence data have been deposited in GenBank (project accession: PRJNA391054).

To identify single nucleotide variants (SNVs) reads were mapped to species-appropriate reference genomes (*E. coli* CFT073 [GenBank: AE014075.1] and *K. pneumoniae* MGH78578 [GenBank: CP000647.1]), and variants called as described previously(30).

316 Alignments of variable sites were padded to the length of the reference genome using bases
317 with the same %GC content as that observed within each dataset. Bootstrapped, maximum-
318 likelihood phylogenies were reconstructed for each species using RaxML version 7.7.6(31),
319 using a generalised time-reversible model and four categories of rate heterogeneity
320 (./RAxML-7.7.6/raxmlHPC-PTHREADS-SSE3 -f a -s <input_alignment.phy> -m
321 GTRGAMMA -p 12345 -c 4 -x 12345 -# 100 -n <output_raxml_rapid_bootstrap>).
322 Phylogenies have been deposited as projects in MicroReact to enable an interactive
323 assessment of geographic distribution of genotypes (*E. coli*:
324 https://microreact.org/project/By8bf5ajg; *K. pneumoniae*:
325 https://microreact.org/project/Hy_yQcaog(32).

326

327 Contigs were assembled using Velvet/VelvetOptimiser (hash value range: 75-149)(33, 34). *In*
328 *silico* MLST was determined by BLASTn(35) matches (100% match) to the Achtman/Pasteur
329 MLST schemes for *E. coli* and *K. pneumoniae*(36, 37), and supported correct species
330 identification. The presence/absence of resistance genes was determined using BLASTn and
331 an in-house curated resistance gene database of over 60 gene families(38). Genes were
332 considered present if a blast match of ≥80% of the query sequence was identified at ≥80%
333 sequence identity using the *de novo* assemblies as blast databases. Ambler class genotype
334 was class A if $bla_{\text{CTX-M}}$, bla_{SHV} and/or bla_{VEB} were present, and/or class C if $bla_{\text{CMY-2}}$, bla_{DHA}
335 and $bla_{\text{ACT-like}}$ genes were present. Where patient faecal samples yielded ≥2 strains, all
336 resistance genes were treated as a single entity within the individual's profile.

337

338 The genetic context of $bla_{\text{CTX-M}}$ was examined by extracting the contigs containing these
339 genes, and annotating these using PROKKA(39), combined with BLASTn and manual

annotation with reference to mobile genetic elements in the ISFinder database(40). Gene locations were characterised as "chromosomal" if other annotations on the contig were only found in chromosomal contexts in the top 20 BLASTn hits when the contig was compared with bacterial sequences available in GenBank (using default parameters); "plasmid" if the other annotations matched only plasmid sequences; or unknown if these conditions were not met e.g. the assembled contigs were too short to verify this.

Epidemiological analyses

Information regarding putative risk factors for ESC-R EC/KP colonisation (collected on a standardised form) included details on: gender, age, hospitalisation status, residence in Siem Reap province versus elsewhere, water source (river, rain, well, bottled, piped, boiled), domestic animals (cats, dogs, birds), livestock (chickens, ducks, pigs, cows or water buffalo), toilet availability, malnutrition, co-morbidities, presence/absence of diarrhoea, presence/absence of parasites (assessed within (28)), soap usage for hand-washing and school attendance. No details regarding antibiotic consumption were ascertained within the study, but previous work locally has shown that individuals are often ill-informed about the nature of any medications used and that 32% of outpatient attendees have evidence of urinary antimicrobial activity(41).

Statistical analyses

Independent risk factors for carriage were identified from a multivariable, stepwise, logistic regression model based on complete cases and initially including all factors (backwards elimination using exit $p<0.1$ to reduce over-fitting). A final multivariable logistic model was then fitted including all cases for which complete information was available for the retained

risk factors. Statistical analyses were performed using STATA version 14 (StataCorp, College Station, USA).

Ethical Approval

The study was approved by the Institutional Review Board (IRB), Angkor Hospital for Children, and the Oxford Tropical Research Ethics Committee (OXTREC 12–12). Caregivers of all included participants gave informed consent for their child to participate in the helminth survey, and for the samples to be used more widely in additional studies approved by the IRB.

374 **Acknowledgements**

375 The authors wish to thank the staff and patients at Angkor Hospital for Children, Siem Reap,

376 Cambodia, and members of the Modernising Medical Microbiology Informatics Group.

377

378 This work was supported by the National Institute for Health Research (NIHR) Oxford

379 Biomedical Research Center (BRC). JvA is currently funded through a National Institute for

380 Health Research (NIHR) Academic Clinical Fellowship. NS is currently funded through a

381 PHE/NIHR/University of Oxford Clinical Lectureship; the sequencing work was also partly

382 funded through a previous Wellcome Trust Doctoral Research Fellowship (#099423/Z/12/Z).

383 TEAP and DWC are NIHR Senior Investigators.

384

385 The funders had no role in study design, data collection and interpretation, or the decision to

386 submit the work for publication. The views expressed are those of the author(s) and not

387 necessarily those of the NHS, the NIHR or the Department of Health.

388

389 The authors have no conflicts of interest to declare.

FIGURE LEGENDS

Figure 1.

Phylogeny of study *Escherichia coli* isolates. Interactive map of geographic locations and genetic attributes can be visualised at: https://microreact.org/project/By8bf5ajg

Figure 2.

Phylogeny of study *Klebsiella pneumoniae* isolates. Interactive map of geographic locations and genetic attributes can be visualised at: https://microreact.org/project/Hy_yQcaog

Figure 3. Schematic of aligned genetic contexts for $bla_{\text{CTX-M-15}}$ in study *Escherichia coli*. Features of interest are highlighted in the figure key. White numbers within open reading frames denote truncated sequence length (bp). Isolates harbouring this genetic context are listed to the left of the figure. "x" denotes contig breaks. ^{P} denotes plasmid contexts; ^{c} chromosomal contexts.

Figure 4. Schematic of aligned genetic contexts for $bla_{\text{CTX-M-55}}$ in study *Escherichia coli*. Features of interest are highlighted in the figure key. White numbers within open reading frames denote truncated sequence length (bp). Isolates harbouring this genetic context are listed to the left of the figure. "x" denotes contig breaks. ^{P} denotes plasmid contexts; ^{c} chromosomal contexts.

Figure 5. Schematic of aligned genetic contexts for $bla_{\text{CTX-M-14}}$ in study *Escherichia coli* and *Klebsiella pneumoniae*. Features of interest are highlighted in the figure key. White numbers within open reading frames denote truncated sequence length (bp). Isolates harbouring this

414 genetic context are listed to the left of the figure. “x” denotes contig breaks. [P] denotes

415 plasmid contexts; [c] chromosomal contexts.

TABLES

Table 1. Clinical and epidemiological details of all 148 participants, also categorised by presence/absence of gastrointestinal colonisation with ESC-resistant *E. coli* and/or *K. pneumoniae*, and multivariable logistic regression outcomes

	Overall (n=148)	ESC-R *E. coli*/ *K. pneumoniae* colonised (n=82; 55%)	ESC-R *E. coli*/ *K. pneumoniae* non-colonised (n=66; 45%)	Univariable logistic regression for ESC-R carriage		Multivariable logistic regression for ESC-R carriage[¥]	
	Number (%) unless otherwise specified	n (%)	n (%)	OR [95% CI]	p	OR [95% CI]	p
Median age [IQR], years	4.24 [1.10 - 8.82]	3.07 [0.97 - 7.21]	6.23 [1.32 - 9.23]	1.00 [1.00-1.00]	0.712		
Male	70 (47)	35 (43)	35 (53)	0.66 [0.34-1.27]	0.211	0.39 [0.18-0.84]	**0.015**
Inpatient[*]	70 (48)	49 (60)	20 (32)	3.28 [1.66-6.50]	**0.001**	3.64 [1.71-7.74]	**0.001**
Province							

Siem Reap	99 (67)	51 (62)	48 (72)	1			
Other (versus Siem Reap)[¶]	49 (33)	31 (38)	18 (28)	1.06 [0.72-1.58]	0.716		
Malnutrition	16 (11)	12 (15)	4 (6)	2.66 [0.81-8.67]	0.105		
Co-morbidities[§]	25 (17)	19 (23)	6 (9)	3.01 [1.13-8.06]	**0.028**		
Diarrhoea present[&]	70 (48)	44 (54)	26 (40)	1.78 [0.92-3.46]	0.086		
Water sources							
Well	123 (83)	64 (78)	59 (89)	0.42 [0.16-1.08]	0.073		
Bottled	17 (11)	13 (16)	4 (6)	2.92 [0.90-9.43]	0.073		
River	5 (3)	3 (4)	2 (3)	1.22 [0.20-7.49]	0.834		
Rain	8 (5)	5 (6)	3 (5)	1.36 [0.31-5.93]	0.680		
School attendance	71 (48)	32 (39)	39 (59)	0.44 [0.23-0.86]	**0.016**	0.39 [0.18-0.83]	**0.015**
All animals	119 (80)	65 (79)	54 (82)				

Domestic animals	113 (76)	61 (74)	50 (79)	0.78 [0.36-1.69]	0.532
Cat	51 (34)	32 (39)	19 (29)	1.58 [0.79-3.17]	0.194
Dog	100 (68)	57 (70)	43 (65)	1.22 [0.61-2.43]	0.573
Birds	24 (16)	12 (15)	12 (18)	0.77 [0.32-1.85]	0.561
Livestock/food animals	89 (60)	54 (66)	35 (53)	1.71 [0.88-3.32	0.114
Water buffalo	4 (3)	3 (4)	1 (1)	2.47 [0.25-24.3]	0.439
Chickens	80 (54)	49 (60)	31 (47)	1.67 [0.87-3.23]	0.122
Pigs	23 (16)	14 (17)	9 (14)	1.30 [0.53-3.23]	0.567
Ducks	2 (1)	2 (2)	0	1 (omitted)	
Cattle	24 (16)	15 (18)	9 (14)	1.42 [0.58-3.48]	0.446

Use of toilet for defecation	88 (59)	45 (54)	42 (65)	0.65 [0.33-1.27]	0.207		
Use of soap[#]							
Never	34 (23)	18 (23)	16 (25)	1			
Some use (versus Never)	111 (77)	62 (76)	49 (75)	1.12 [0.52-2.43]	0.765		
Presence of intestinal parasites	36 (24)	25 (30.49)	11 (17.19)	2.19 [0.99-4.88]	0.054	3.96 [1.55-10.12]	**0.004**

[*] missing one datapoint (n=147)

[&] missing two datapoints (n=146)

[#] missing three datapoints (n=145)

[¥] Backwards elimination performed using 144 cases for which complete information available, on all predictors, using exit p≤0.1. Final model then incorporated 147 cases for which complete information available on included predictors (gender, in-patient status, presence of intestinal parasites and school attendance)

[¶] "Other province" category includes: Banteay Meanchey (n=21), Oddar Meanchey (8), Kampong Thom (6), Battambang (5), Preah Vihear (3), Kampong Cham (2), Kampong Chhang (1), Lann Kimfong (1), Pursat (1), Vuth Variath (1)

[§] Includes: HIV (n=5), blood dyscrasia (3), Down's syndrome (2), congenital heart disease (6), TB (4), asthma (2), other (8); five individuals had multiple co-morbidities

p-values <0.05 in bold

419 **Table 2. Summary of genetic contexts of *bla*$_{CTX-M}$ in *E. coli***

	E. coli bla$_{CTX-M-15}$ (n=41)	*E. coli bla*$_{CTX-M-55}$ (n=24)	*E. coli bla*$_{CTX-M-14}$ (n=15)	*E. coli bla*$_{CTX-M-27}$ (n=12)	*E. coli bla*$_{CTX-M-24}$ (n=3)
Location (chromosome versus plasmid)					
Chromosomal	5 (12%)	4 (17%)	2 (13%)	0	2 (67%)
Likely plasmid	2 (5%)	4 (17%)	5 (33%)	0	1 (33%)
Not determined	34 (83%)	16 (67%)	8 (53%)	12 (100%)	0
Evaluable immediate flanking context	**n=40***	**n=23¥**	**n=15**	**n=12**	**n=3**
IS*Ecp1* upstream	**40 (100%)**	**23 (100%)**	**15 (100%)**	**12 (100%)**	**3 (100%)**
3' GACTA target site sequence (TSS) at	36 (90%)	23 (100%)	0	0	0

48bp upstream of *bla*$_{CTX-M}$					
3’ TTTCA TSS at 127bp upstream of *bla*$_{CTX-M}$	4 (10%)	0	0	0	0
3’ GAATA TSS at 42bp	0	0	15 (100%)	12 (100%)	3 (100%)
IS*Ecp1* incomplete	**26 (65%)**	**14 (61%)**	**6 (40%)**	**12 (100%)**	**1 (100%)**
due to IS*26* element	13** (32%) left IRR - 11 right IRR - 2	12† (52%) left IRR - 1 right IRR - 11	0	12 (100%) left IRR - 12	0
due to contig break	13 (32%)	2 (9%)	3 (20%)	0	1 (33%)
due to IS*Vsa5*-like	0	0	2 (13%)	0	0

due IS*1S* right IRR	0	0	1 (7%)	0	0
IS*Ecp1* complete	**13 (32%)**	**9 (39%)**	**9†† (60%)**	**0**	**2 (67%)**
5' TCATA TSS	9 (22%)	4 (17%)	0	0	0
5' TAATA TSS	4 (10%)	3 (13%)	0	0	0
5' TAACA TSS	0	2 (14%)	0	0	0
5' CATTA TSS	0	0	4 (44%)	0	1 (33%)
5' AAATA TSS	0	0	2 (22%)	0	0
5' TAAAA TSS	0	0	1 (11%)	0	0
5' GCCGA TSS	0	0	1 (11%)	0	0
5' TATAT TSS	0	0	0	0	1 (33%)

420 * excluding one isolate with short contigs harbouring truncated $bla_{\text{CTX-M-15}}$

421 ** IS*Ecp1* truncated at 24, 497, 524, 1067, 1173, 1421, or 1489bp.

422 ¥ excluding one isolate with short contig harbouring truncated $bla_{\text{CTX-M-55}}$

423 † IS*Ecp1* truncated at 267, 309 or 497bp

†† For one only 1 bp of 5’ TSS evaluable

¶ IS*Ecp1* truncated at 149, 192, 208 and 388bp

TSS=target site sequence

IRR=inverted repeat region

REFERENCES

1. **Carlet J.** 2012. The gut is the epicentre of antibiotic resistance. Antimicrob Resist Infect Control **1:**39.

2. **Selden R, Lee S, Wang WL, Bennett JV, Eickhoff TC.** 1971. Nosocomial klebsiella infections: intestinal colonization as a reservoir. Ann Intern Med **74:**657-664.

3. **Pitout JD, Laupland KB.** 2008. Extended-spectrum beta-lactamase-producing Enterobacteriaceae: an emerging public-health concern. Lancet Infect Dis **8:**159-166.

4. **Rottier WC, Ammerlaan HS, Bonten MJ.** 2012. Effects of confounders and intermediates on the association of bacteraemia caused by extended-spectrum beta-lactamase-producing Enterobacteriaceae and patient outcome: a meta-analysis. J Antimicrob Chemother **67:**1311-1320.

5. **Schwaber MJ, Navon-Venezia S, Kaye KS, Ben-Ami R, Schwartz D, Carmeli Y.** 2006. Clinical and economic impact of bacteremia with extended- spectrum-beta-lactamase-producing Enterobacteriaceae. Antimicrob Agents Chemother **50:**1257-1262.

6. **Luvsansharav UO, Hirai I, Nakata A, Imura K, Yamauchi K, Niki M, Komalamisra C, Kusolsuk T, Yamamoto Y.** 2012. Prevalence of and risk factors associated with faecal carriage of CTX-M beta-lactamase-producing Enterobacteriaceae in rural Thai communities. J Antimicrob Chemother **67:**1769-1774.

7. **Stoesser N, Crook DW, Moore CE, Phetsouvanh R, Chansamouth V, Newton PN, Jones N.** 2012. Characteristics of CTX-M ESBL-producing Escherichia coli isolates from the Lao People's Democratic Republic, 2004-09. J Antimicrob Chemother **67:**240-242.

8. **Woerther PL, Burdet C, Chachaty E, Andremont A.** 2013. Trends in human fecal carriage of extended-spectrum beta-lactamases in the community: toward the globalization of CTX-M. Clin Microbiol Rev **26:**744-758.

9. **Vlieghe ER, Huang TD, Phe T, Bogaerts P, Berhin C, De Smet B, Peetermans WE, Jacobs JA, Glupczynski Y.** 2015. Prevalence and distribution of beta-lactamase coding genes in third-generation cephalosporin-resistant Enterobacteriaceae from bloodstream infections in Cambodia. Eur J Clin Microbiol Infect Dis **34:**1223-1229.

10. **Ruppe E, Hem S, Lath S, Gautier V, Ariey F, Sarthou JL, Monchy D, Arlet G.** 2009. CTX-M beta-lactamases in Escherichia coli from community-acquired urinary tract infections, Cambodia. Emerg Infect Dis **15:**741-748.

11. **Moore CE, Sona S, Poda S, Putchhat H, Kumar V, Sopheary S, Stoesser N, Bousfield R, Day N, Parry CM.** 2016. Antimicrobial susceptibility of uropathogens isolated from Cambodian children. Paediatr Int Child Health doi:10.1080/20469047.2015.1109241**:**1-5.

12. **Turner P, Pol S, Soeng S, Sar P, Neou L, Chea P, Day NP, Cooper BS, Turner C.** 2016. High Prevalence of Antimicrobial-resistant Gram-negative Colonization in Hospitalized Cambodian Infants. Pediatr Infect Dis J **35:**856-861.

13. **Woodford N, Turton JF, Livermore DM.** 2011. Multiresistant Gram-negative bacteria: the role of high-risk clones in the dissemination of antibiotic resistance. FEMS Microbiol Rev **35:**736-755.

14. **Stoesser N, Emary K, Soklin S, Peng An K, Sophal S, Chhomrath S, Day NP, Limmathurotsakul D, Nget P, Pangnarith Y, Sona S, Kumar V, Moore CE, Chanpheaktra N, Parry CM.** 2013. The value of intermittent point-prevalence surveys of healthcare-associated infections for evaluating infection control

477 interventions at Angkor Hospital for Children, Siem Reap, Cambodia. Trans R Soc
478 Trop Med Hyg **107:**248-253.
479 15. **Kurz MS, Bayingana C, Ndoli JM, Sendegeya A, Durst A, Pfuller R, Gahutu JB,**
480 **Mockenhaupt FP.** 2017. Intense pre-admission carriage and further acquisition of
481 ESBL-producing Enterobacteriaceae among patients and their caregivers in a tertiary
482 hospital in Rwanda. Trop Med Int Health **22:**210-220.
483 16. **Lee SC, Tang MS, Lim YA, Choy SH, Kurtz ZD, Cox LM, Gundra UM, Cho I,**
484 **Bonneau R, Blaser MJ, Chua KH, Loke P.** 2014. Helminth colonization is
485 associated with increased diversity of the gut microbiota. PLoS Negl Trop Dis
486 **8:**e2880.
487 17. **Fernandez-Reyes M, Vicente D, Gomariz M, Esnal O, Landa J, Onate E, Perez-**
488 **Trallero E.** 2014. High rate of fecal carriage of extended-spectrum-beta-lactamase-
489 producing Escherichia coli in healthy children in Gipuzkoa, northern Spain.
490 Antimicrob Agents Chemother **58:**1822-1824.
491 18. **Ling Oh A, Hassali MA, Al-Haddad MS, Syed Sulaiman SA, Shafie AA, Awaisu**
492 **A.** 2011. Public knowledge and attitudes towards antibiotic usage: a cross-sectional
493 study among the general public in the state of Penang, Malaysia. J Infect Dev Ctries
494 **5:**338-347.
495 19. **Gebeyehu E, Bantie L, Azage M.** 2015. Inappropriate Use of Antibiotics and Its
496 Associated Factors among Urban and Rural Communities of Bahir Dar City
497 Administration, Northwest Ethiopia. PLoS One **10:**e0138179.
498 20. **Ben-Ami R, Rodriguez-Bano J, Arslan H, Pitout JD, Quentin C, Calbo ES, Azap**
499 **OK, Arpin C, Pascual A, Livermore DM, Garau J, Carmeli Y.** 2009. A
500 multinational survey of risk factors for infection with extended-spectrum beta-

lactamase-producing enterobacteriaceae in nonhospitalized patients. Clin Infect Dis **49:**682-690.

21. **Pena C, Gudiol C, Tubau F, Saballs M, Pujol M, Dominguez MA, Calatayud L, Ariza J, Gudiol F.** 2006. Risk-factors for acquisition of extended-spectrum beta-lactamase-producing Escherichia coli among hospitalised patients. Clin Microbiol Infect **12:**279-284.

22. **Coelho A, Gonzalez-Lopez JJ, Miro E, Alonso-Tarres C, Mirelis B, Larrosa MN, Bartolome RM, Andreu A, Navarro F, Johnson JR, Prats G.** 2010. Characterisation of the CTX-M-15-encoding gene in Klebsiella pneumoniae strains from the Barcelona metropolitan area: plasmid diversity and chromosomal integration. Int J Antimicrob Agents **36:**73-78.

23. **He S, Hickman AB, Varani AM, Siguier P, Chandler M, Dekker JP, Dyda F.** 2015. Insertion Sequence IS26 Reorganizes Plasmids in Clinically Isolated Multidrug-Resistant Bacteria by Replicative Transposition. MBio **6:**e00762.

24. **Partridge SR, Zong Z, Iredell JR.** 2011. Recombination in IS26 and Tn2 in the evolution of multiresistance regions carrying blaCTX-M-15 on conjugative IncF plasmids from Escherichia coli. Antimicrob Agents Chemother **55:**4971-4978.

25. **Stoesser N, Sheppard AE, Pankhurst L, De Maio N, Moore CE, Sebra R, Turner P, Anson LW, Kasarskis A, Batty EM, Kos V, Wilson DJ, Phetsouvanh R, Wyllie D, Sokurenko E, Manges AR, Johnson TJ, Price LB, Peto TE, Johnson JR, Didelot X, Walker AS, Crook DW, Modernizing Medical Microbiology Informatics G.** 2016. Evolutionary History of the Global Emergence of the Escherichia coli Epidemic Clone ST131. MBio **7:**e02162.

26. **Stoesser N, Sheppard AE, Moore CE, Golubchik T, Parry CM, Nget P, Saroeun M, Day NP, Giess A, Johnson JR, Peto TE, Crook DW, Walker AS, Modernizing**

526 **Medical Microbiology Informatics G.** 2015. Extensive Within-Host Diversity in
527 Fecally Carried Extended-Spectrum-Beta-Lactamase-Producing Escherichia coli
528 Isolates: Implications for Transmission Analyses. J Clin Microbiol **53:**2122-2131.
529 27. **Arcilla MS, van Hattem JM, Haverkate MR, Bootsma MC, van Genderen PJ,**
530 **Goorhuis A, Grobusch MP, Lashof AM, Molhoek N, Schultsz C, Stobberingh**
531 **EE, Verbrugh HA, de Jong MD, Melles DC, Penders J.** 2017. Import and spread of
532 extended-spectrum beta-lactamase-producing Enterobacteriaceae by international
533 travellers (COMBAT study): a prospective, multicentre cohort study. Lancet Infect
534 Dis **17:**78-85.
535 28. **Moore CE, Nget P, Saroeun M, Kuong S, Chanthou S, Kumar V, Bousfield R,**
536 **Nader J, Bailey JW, Beeching NJ, Day NP, Parry CM.** 2015. Intestinal parasite
537 infections in symptomatic children attending hospital in Siem Reap, Cambodia. PLoS
538 One **10:**e0123719.
539 29. **Chemotherapy BSfA.** 2012. Detection of extended-spectrum beta-lactamases
540 (ESBLs) in E. coli and Klebsiella species. http://bsac.org.uk/wp-
541 content/uploads/2012/02/Ecoliklebsiella.pdf. Accessed
542 30. **Stoesser N, Xayaheuang S, Vongsouvath M, Phommasone K, Elliott I, Del Ojo**
543 **Elias C, Crook DW, Newton PN, Buisson Y, Lee SJ, Dance DA.** 2015.
544 Colonization with Enterobacteriaceae producing ESBLs in children attending pre-
545 school childcare facilities in the Lao People's Democratic Republic. J Antimicrob
546 Chemother **70:**1893-1897.
547 31. **Stamatakis A.** 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic
548 analyses with thousands of taxa and mixed models. Bioinformatics **22:**2688-2690.
549 32. **Aanensen DM, Feil EJ, Holden MT, Dordel J, Yeats CA, Fedosejev A, Goater R,**
550 **Castillo-Ramirez S, Corander J, Colijn C, Chlebowicz MA, Schouls L, Heck M,**

Pluister G, Ruimy R, Kahlmeter G, Ahman J, Matuschek E, Friedrich AW, Parkhill J, Bentley SD, Spratt BG, Grundmann H, European SRLWG. 2016. Whole-Genome Sequencing for Routine Pathogen Surveillance in Public Health: a Population Snapshot of Invasive Staphylococcus aureus in Europe. MBio **7**.

33. **Zerbino DR.** 2010. Using the Velvet de novo assembler for short-read sequencing technologies. Curr Protoc Bioinformatics **Chapter 11:**Unit 11 15.

34. **Gladman SS, T.** VelvetOptimizer. http://bioinformatics.net.au/software.velvetoptimiser.shtml. Accessed

35. **Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ.** 1990. Basic local alignment search tool. J Mol Biol **215:**403-410.

36. **Diancourt L, Passet V, Verhoef J, Grimont PA, Brisse S.** 2005. Multilocus sequence typing of Klebsiella pneumoniae nosocomial isolates. J Clin Microbiol **43:**4178-4182.

37. **Wirth T, Falush D, Lan R, Colles F, Mensa P, Wieler LH, Karch H, Reeves PR, Maiden MC, Ochman H, Achtman M.** 2006. Sex and virulence in Escherichia coli: an evolutionary perspective. Mol Microbiol **60:**1136-1151.

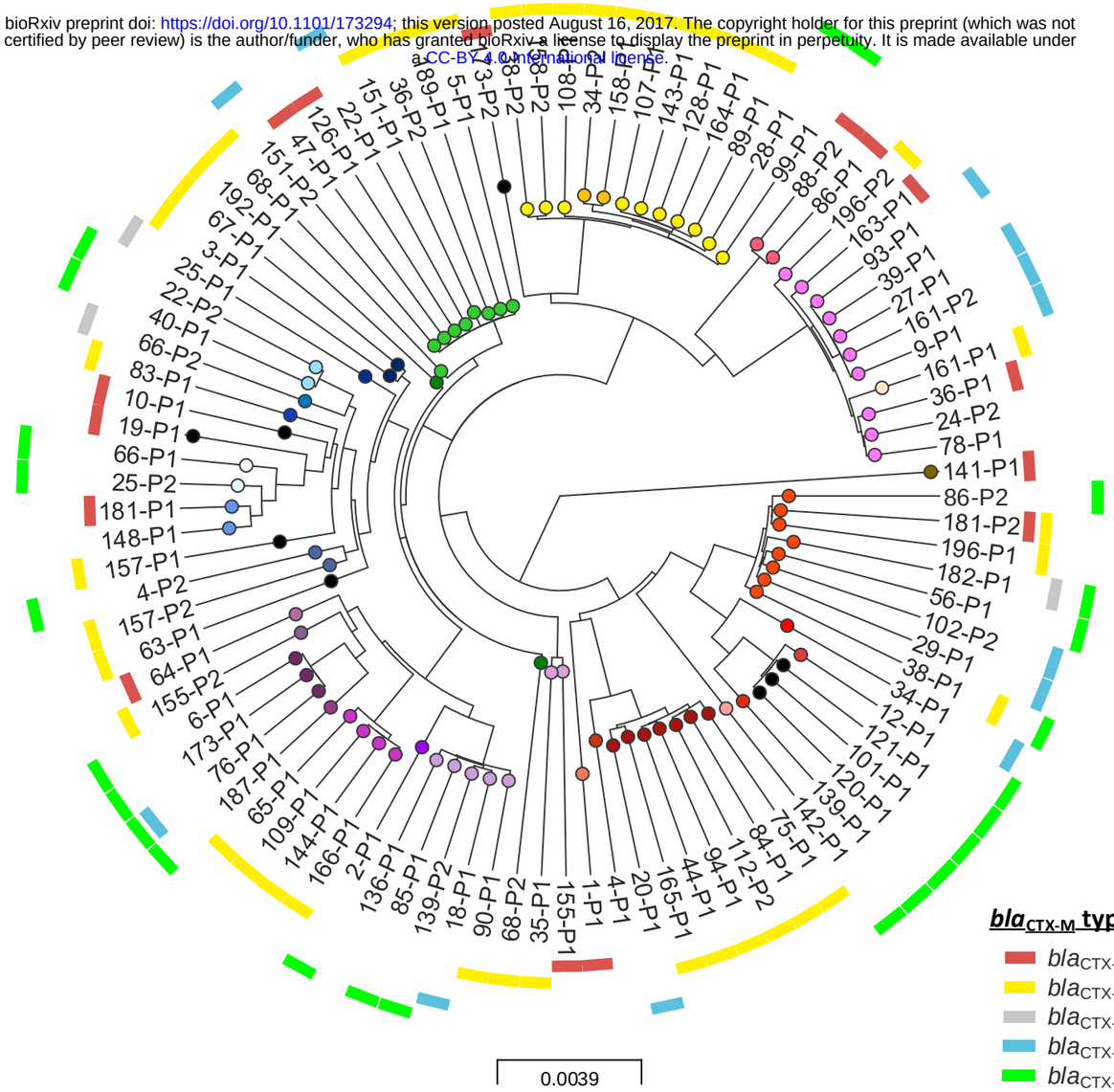
38. **Stoesser N, Batty EM, Eyre DW, Morgan M, Wyllie DH, Del Ojo Elias C, Johnson JR, Walker AS, Peto TE, Crook DW.** 2013. Predicting antimicrobial susceptibilities for Escherichia coli and Klebsiella pneumoniae isolates using whole genomic sequence data. J Antimicrob Chemother **68:**2234-2244.

39. **Seemann T.** 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics **30:**2068-2069.

40. **Siguier P, Perochon J, Lestrade L, Mahillon J, Chandler M.** 2006. ISfinder: the reference centre for bacterial insertion sequences. Nucleic Acids Res **34:**D32-36.

575 41. **Emary KR, Carter MJ, Pol S, Sona S, Kumar V, Day NP, Parry CM, Moore CE.**
576 2015. Urinary antibiotic activity in paediatric patients attending an outpatient
577 department in north-western Cambodia. Trop Med Int Health **20:**24-28.

578

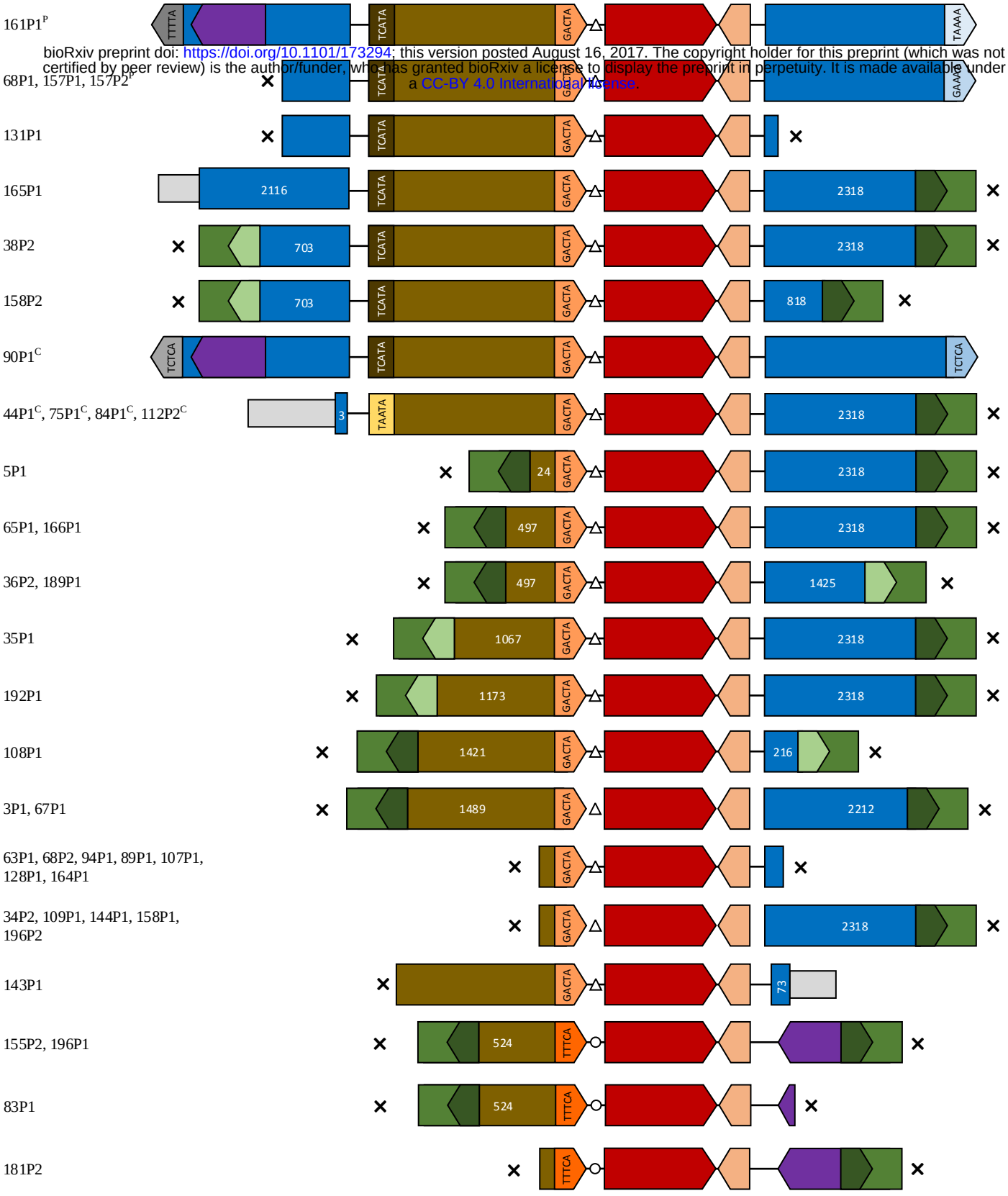


MSLT

- Novel ST
- ST354
- ST1722
- ST1674
- ST457
- ST3032
- ST648
- ST1458
- ST2914
- ST617
- ST156
- ST155
- ST410
- ST101
- ST196
- ST469
- ST1136
- ST48
- ST46
- ST609
- ST4539
- ST4429
- ST226
- ST2345
- ST6143
- ST167
- ST44
- ST131
- ST405
- ST5954
- ST394
- ST38
- ST3268
- ST6714

bla_{CTX-M} type

- $bla_{CTX-M-14}$
- $bla_{CTX-M-15}$
- $bla_{CTX-M-24}$
- $bla_{CTX-M-27}$
- $bla_{CTX-M-55}$



Key

L IRR — IS*26* — R IRR : IS*26*

Unknown : Unknown flanking sequence

bla_{TEM-1} : bla_{TEM-1}

127bp, 48bp : Sequence upstream of $bla_{CTX-M-15}$

L TSS — **Tn2 – $tnpR_{Tn2}$ and bla_{TEM}** — : Tn2 flanking sequence and target site sequences (TSS)

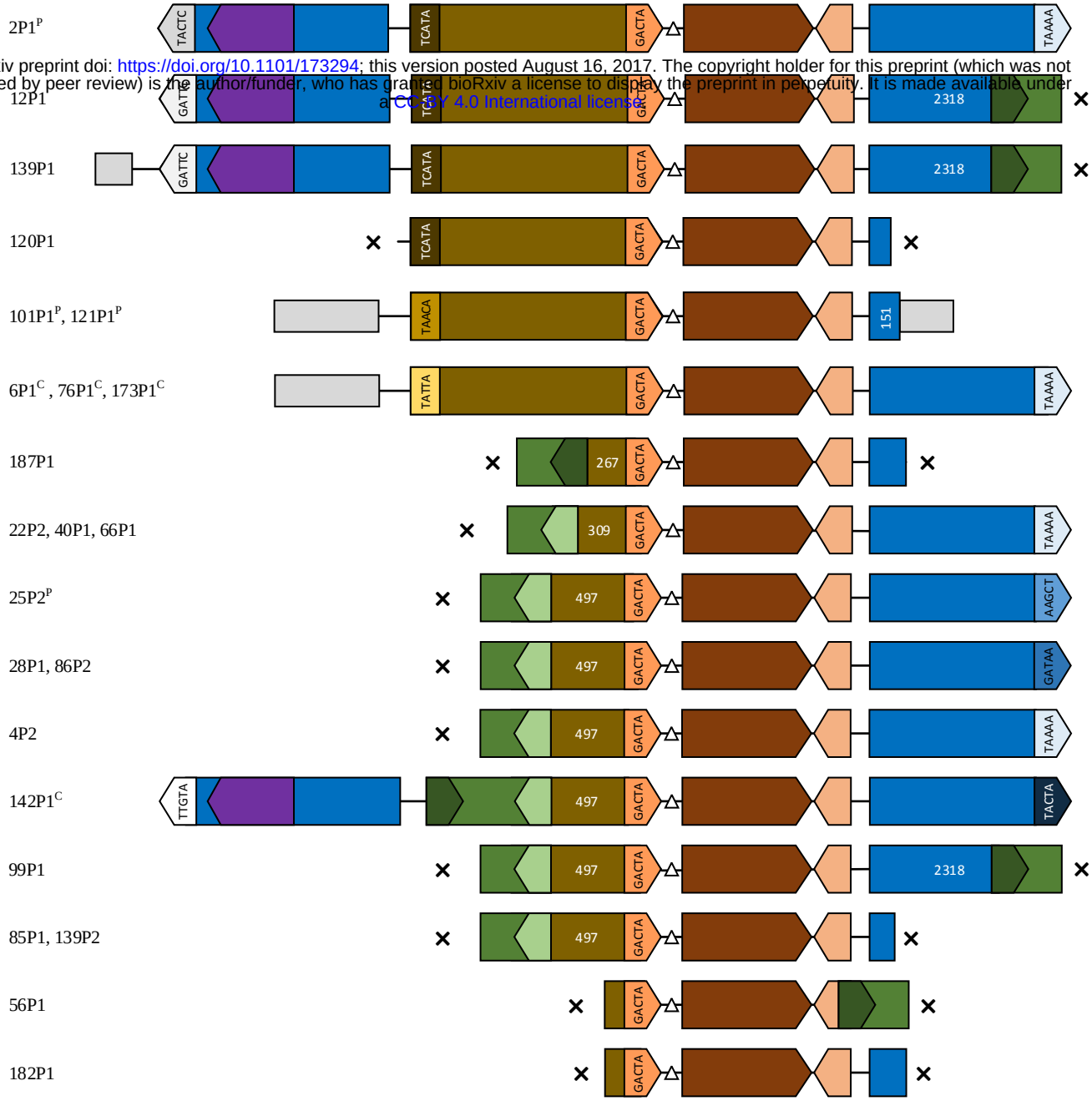
Tn2 – $tnpA_{Tn2}$ — R TSS

TCTCA, TTTTA, TAAAA, GAAAA, TCTCA

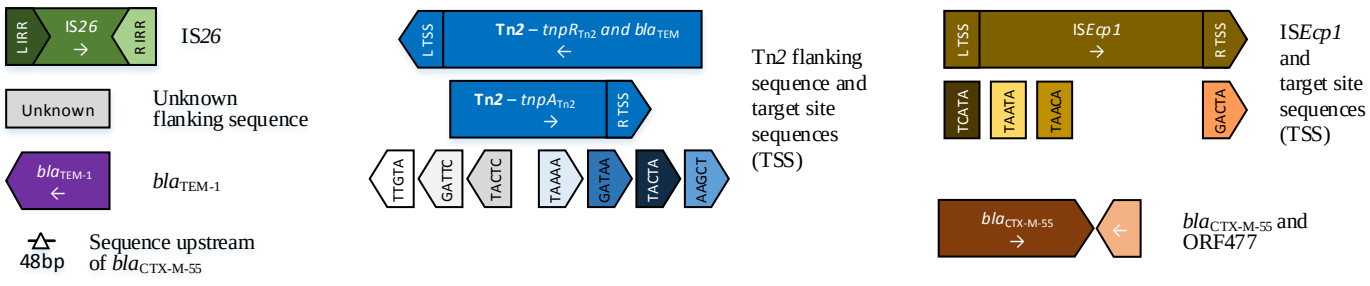
L TSS — IS*Ecp1* — R TSS : IS*Ecp1* and target site sequences (TSS)

TCATA, TAATA, TTTCA, GACTA

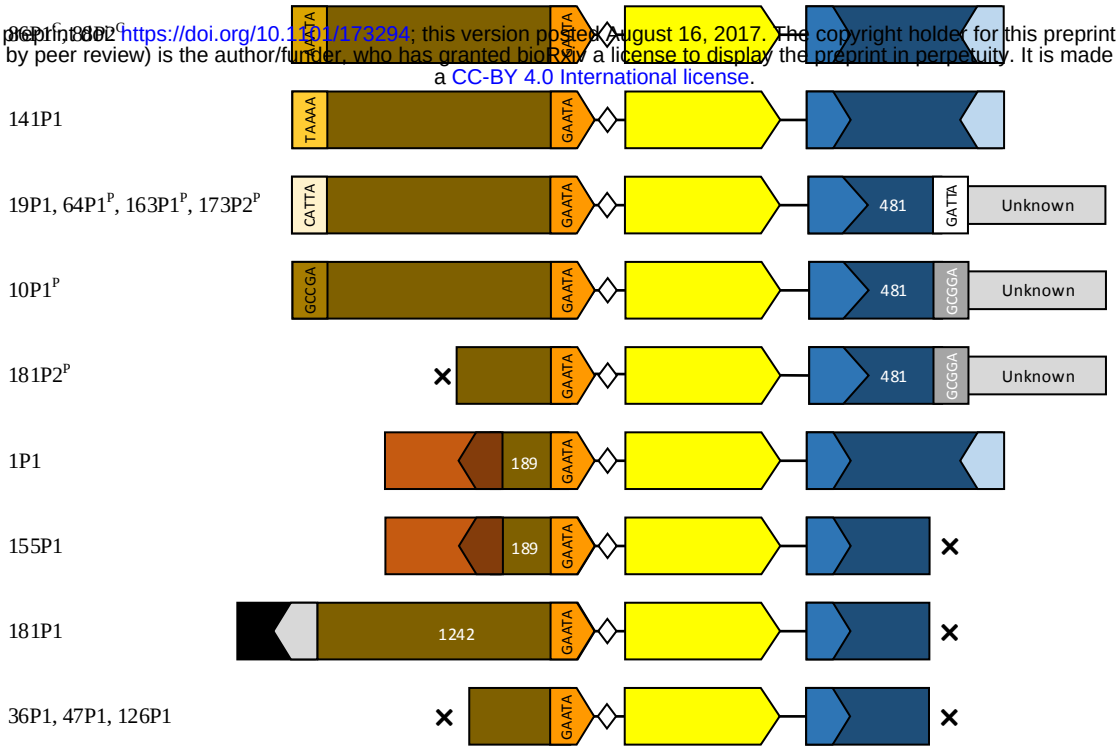
$bla_{CTX-M-15}$: $bla_{CTX-M-15}$ and ORF477



Key



a) ***E. coli***



b) ***K. pneumoniae***