Analysis of *Escherichia coli* sequence types and resistance mechanisms in sewage from Islamabad, Pakistan indicates difference in *E. coli* carriage types between S. Asia and Europe

Rabaab ZAHRA[1], Saba JAVEED[1], Bibi MALALA[1], Dmitriy BABENKO[2] and Mark A. TOLEMAN[*3]

Department of Microbiology, Quaid-i-Azam University, Islamabad, Pakistan[1]. Karaganda State Medical University, Karaganda, Kazakhstan[2]. Department of Infection and immunity, Cardiff University, UK[3].

[*]Corresponding author: Dr Mark A. Toleman, Department of Infection and immunity, Cardiff University, UK.

Running title: Carriage *E. coli* sequence types in Islamabad, Pakistan.

Abstract:

Objectives: To discover the *E. coli* carriage sequence types and associated resistance mechanisms in the community in Islamabad, Pakistan by analysis of *E. coli* isolates in sewage.

Methods: 110 *E. coli* were isolated from sewage across the city of Islamabad without antibiotic bias and confirmed as *E. coli* by MALDI-TOF. Isolates were characterized by CH typing and cgMLST. Resistance mechanisms, virulence genes, phylotypes and plasmid incompatibility types were determined in a subset of isolates by in-silico analysis. The genomic position of *bla*$_{\text{CTX-M-15}}$ was determined using S1 PFGE, probing and Nanopore minion sequencing.

Results and conclusions: The most prevalent sequence types were ST394, ST10 and ST648 accounting for 39% of all isolates collected and were found at many sites across Islamabad. Carbapenemase genes were absent and only a single isolate of ST131 was found. *qnr*S1 and *bla*$_{\text{CTX-M-15}}$ were the most prevalent resistance mechanisms with *bla*$_{\text{CTX-M-15}}$ penetrating many sequence types and 31% of all collected isolates. However, the majority of the successful sequence types were *bla*$_{\text{CTX-M-15}}$ negative indicating that resistance is not the main driver of prevalence. 23% of *bla*$_{\text{CTX-M-15}}$ genes were chromosomally encoded and large IS*EcP1* mediated insertions included *qnr*S1 and several plasmid genes. In all chromosomally encoded isolates no plasmid copies of *bla*$_{\text{CTX-M-15}}$ were found. The most prevalent ST (ST394) contained many Enteroaggregative *E. coli* (EAEC) virulence genes and the *fim*H30 variant allele previously linked to the success of ST131.

Introduction:

Antibiotic resistance in *Escherichia coli* is a major concern as *E. coli* is both a leading cause of human infection and an ever-present gut colonist. Carriage of ESBL *E. coli* can lead to difficult to treat urinary tract infections (UTIs), delay appropriate therapy and lead to poor outcomes.[1]

The *E. coli* species consists of many different sequence types (>7,000 http://pubmlst.org/databases.shtml) and strains.[2] Individual strains vary in ability to cause disease and to resist antibiotics. Several molecular methodologies have enabled comparisons of strains between diverse geographic locations including: MLST; *fim*H/*fum*C (CH typing); [3] Phylo-typing [4] and WGS. Comparisons have revealed that resistance is associated with a small number of successful strains and sequence types, for example ST131 *E. coli* belonging to phylogroup B2 [5] and *E. coli* ST101 belonging to the phylogroup B1.[6] We have also seen expansion of ST131 ESBL *E. coli* causing severe infections.[7] A recent study investigating ESBL *E. coli* in the UK found ST131 was the primary cause of cephalosporin resistant bacteraemia's and the most prevalent ST in faeces and sewage (H. Cadden, P. Cleary, M. Day, M. Doumith, M. Ellington, N. Elviss, J. Findlay, K. Hopkins, B. Jones, D. Livermore, L. Randall, C. Teale, M. Toleman, D. Wareham, C. Wiuff, N. Woodford unpublished results). This suggests that analysis of *E. coli* in sewage is a good proxy for gut carriage in the community.

In this study, we sought to determine the carriage *E. coli* sequence types/resistance mechanisms by analysis of *E. coli* collected without antibiotic selection from 18 sewage outfall sites across the city of Islamabad.

Materials and Methods:

Collection of samples: Samples (30 mL) were collected from 18 sites in 9 sectors of Islamabad (Figure S1).

Bacterial isolation and identification: Samples were centrifuged, pelleted and re-suspended in 0.5mL LB broth. 50 µL was spread on MacConkey plates without antibiotic and grown at 37°C overnight. Ten colonies with typical *E. coli* morphology were randomly collected from each site with a total of 110 colonies confirmed as *E. coli* by MALDI-TOF.

Detection of *bla*$_{\text{CTX-M}}$ and *bla*$_{\text{NDM-1}}$ genes: PCR utilized custom primers and ReddyMix Extensor PCR Master Mix 1 (Thermo Scientific) with appropriate controls.

Two Locus CH typing: *fimH* and *fumC* genes were amplified by PCR as described by Weissman.[3] *fumC* and *fimH* alleles were assigned using (http://mlst.warwick.ac.uk/mlst/dbs/Ecoli) and (https://cge.cbs.dtu.dk/services/FimTyper/) websites, respectively.

***bla*$_{\text{CTX-M-15}}$ genomic location:** S1 PFGE was performed as described.[8] The probe was amplified using primers CTXMF/R (GTTCACGCTGATGGCGACGGC, ACGCTAATACATCGCGACGGC) and radio-labeled using P^{32} dCTP as described previously.[8]

Genomic DNA extraction: Genomic DNA was extracted using the Qiagen genomic DNA kit.

Miseq Sequencing: DNA libraries were prepared using the NexteraXT sample kit and sequenced (20-30× coverage) with a standard 2x100 base protocol on a MiSeq Instrument (Illumina, San Diego, CA, USA) in house at Cardiff.

Sequence type and resistance/virulence gene detection: MLST was determined with StringMLST using short read data in fastq format and Ridom SeqSphere+ (version 3.5.0) using assembled data in fasta format. *E. coli* strains were clustered based on cgMLST typing with SeqSphere+. A core genome MLST (cgMLST) scheme was based on *E. coli* ATCC25922 with 1907 targets. Antimicrobial resistance genes were detected using CLC-Biogenomic workbench. Plasmid and virulence genes were detected using Plasmid finder (https://cge.cbs.dtu.dk/services/PlasmidFinder/); and Virulence finder (https://cge.cbs.dtu.dk/services/VirulenceFinder/).

Minion sequencing: Isolates with chromosomally encoded *bla*$_{\text{CTX-M-15}}$ were sequenced using the nanopore minion RAD-002 rapid sequencing kit following isolation of high molecular weight DNA by the CTAB (hexadecyltrimethylammonium bromide) method.[9] Single reads were used as scaffold to assemble miseq data and annotated using geneious.

Phylo-group analysis: The *E. coli* phylogroups were determined with in-silico searches for *chuA, yjaA, tspE4.C2, arpAgpE, tnpAgpC* using geneious software based on the Clermont method.[4]

Results and discussion

This study was designed to discover *E. coli* carriage ST and resistance mechanisms in Islamabad. We collected *E. coli* at sewage outfalls across the city and isolated 110 *E. coli* strains without antibiotic selection. PCR indicated 34 isolates carried $bla_{\text{CTX-M-15}}$ and a single isolate carried $bla_{\text{CTX-M-27}}$. All isolates were $bla_{\text{NDM-1}}$ negative and carbapenem susceptible indicating low carbapenemase carriage rates in Islamabad as compared to other S. Asian sites. [10 11 12 13]

CH clonotyping revealed 24 known MLST in the collection of 110 *E. coli's*. These were *fum*C- *fim*H: 35-30 (ST394); 11-54 (ST10); 4-0 (ST648); 65-32/ 54-0 (ST1431) and 7-54 (ST45) which accounted for 15%, 13%, 7%, 3.6% and 2.7% of isolates respectively. Two isolates of clonotypes, *fum*C- *fim*H: 4-27 (ST88); 29-38 (ST156); 6-0 (ST688); 7-54 (ST2325); 636-34 (ST5176) and 11-24 (ST43) which equates to 1% prevalence each and single isolates of each clonotype including: 11-27 (ST93); 40-30 (ST131); 6-31 (ST154); 4-32 (ST155); 23-54 (ST205); 11-69 (ST216); 4-61 (ST224); 7-25 (ST398); 136-331 (ST543); 65-32 (ST1128) and 27-0 (ST4121). Details of location, numbers of ST and respective clonotypes are shown in Figure S1. Of 110 *E. coli* isolates, 33 were whole genome sequenced (WGS) including examples of all MLST sequence types and multiple examples of common ones (Table S1). The in silico MLST agreed 100% with MLST derived from CH clonotyping. Most frequently found sequence types were ST394, ST10, ST648 and ST1431 found at 44%, 44%, 28% and 11% of the sampling sites, respectively (Figure S1). Interestingly, ST131 was particularly rare (a single isolate <1%). Isolation of *E. coli* without antibiotic selection enabled us to measure penetration of resistance mechanisms through this species. $bla_{\text{CTX-M-15}}$ is highly prevalent worldwide [14] and colonized 23% of *E. coli*

isolates in Islamabad including 63% of MLST. However, it was found only in 14%, 24% and 50% of individual isolates of the most prevalent types: ST648; ST394 and ST10, respectively. This indicates that *bla*$_{\text{CTX-M-15}}$ *E. coli* are overall less fit than parental isolates and that the prevalence of individual *E. coli* ST is not directly related to cephalosporin resistance.

WGS revealed all resistance and virulence mechanisms in 33 isolates including 24 positive and 9 negative CTX-M isolates (Table S1). The *qnr*S1 gene was most common, found in 80 isolates by PCR. This was often associated with *bla*$_{\text{CTX-M-15}}$ (91%). Other prevalent mechanisms were: *dhfr* genes found in 60%, *bla*$_{\text{TEM-1B}}$ in 33%, *tet*A in 42%, *sul2* in 33%, *sul1* 21% and *tetB* 9% of WGS isolates. Other mechanisms were rarely found: *bla*$_{\text{OXA-1}}$ in 3; *bla*$_{\text{TEM-199}}$ in 2; and *bla*$_{\text{TEM-158}}$, *bla*$_{\text{CMY-42}}$, and *bla*$_{\text{CMY-44}}$ and in individual isolates Table S1.

WGS detected 19 different virulence genes (Table S1). Overall the majority of isolates were commensals with few virulence factors belonging to phylogroups A, C and B1. The ST394 isolates belonged to phylogroup E and harboured entero-aggregative virulence factors typical of the EAEC group of *E. coli*. ST394 is associated with diarrheal disease but also commonly recovered from healthy people and has been implicated in acute and persistent sporadic diarrhea, and outbreaks, in both industrialized and developing countries.[15] ST131 and ST648 were the only EXPEC pathogens recovered belonging to phylogroups B2 and D, respectively. They are known to cause UTI's and sepsis, though they do not carry the classical EXPEC virulence gene repertoire.[16]

S1 PFGE analysis revealed 1-4 plasmids present in each strain. Plasmids were of a range of incompatibility groups (Table S1). The *bla*$_{\text{CTX-M-15}}$ plasmids ranged in size from 50-150 kb. We found *bla*$_{\text{CTX-M-15}}$ genes on the chromosome in 8 isolates

(23%) and 6 ST. We determined the insertion sites in 4 isolates including identical sites in ST394 and a unique insertion in ST4121 (Figure S2). In all cases movement to the chromosome was *ISEcP1* mediated and confirmed by the presence of direct repeats generated at the site of insertion (Figure S2). Notably these movement events not only transposed *bla*$_{\text{CTX-M-15}}$ to the chromosome but also sections of a common plasmid containing *qnr*S1. Movement to the chromosome is typically associated with subsequent loss of the *bla*$_{\text{CTX-M-15}}$ plasmid as evidenced by lack of *bla*$_{\text{CTX-M-15}}$ harboring plasmids in these strains. This process likely enhances the fitness of the strain overall.[17]

We used cgMLST to relate the various *E. coli* to each other (Figure 1). ST were correctly grouped together. The Clermount PCR phylogroup analysis gave good agreement in general with cgMLST. Interestingly, cgMLST aligned the ST648 isolates as group F instead of D, confirming the recent observation of Johnson[18] and highlighting the superior discriminatory power of cgMLST.

Analysis of the *fim*H data revealed that several strains were *fim*H null. This included all ST648 isolates and is typical of this ST.[6] Its prevalence as an EXPEC pathogen suggests that another adhesion is substituting for *fim*H. The worldwide success of the *bla*$_{\text{CTX-M-15}}$ gene in ST131 *E. coli* has been closely linked to a highly adherent *fim*H variant allele, *fim*H30.[5] In this study despite ST131 being particularly rare, *fim*H30 was the second most prevalent variant of *fim*H and was present in all ST394 isolates. The possession of *fim*H30 by ST394 is likely advantageous. However, the fact that ST131 *fim*H30 *E. coli* is particularly rare suggests that other as yet unknown factors are responsible for the success of ST394 in Islamabad.

Acknowledgements: This study was presented as a poster at the British Society of Antimicrobial Chemotherapy Spring meeting in March 2017 "Association of Antibiotic genes with Successful ST of *Escherichia coli* in The Environment."

Funding: This project was funded by a Commonwealth Academic Staff fellowship grant (PKCF-2016-156) awarded to Dr Rabaab Zahra.

Transparency Declaration: All authors have no conflicting interests to declare.

Figure and Table legends:

Figure 1. cgMLST of Islamabad sewage isolates.

cgMLST was generated using a core genome of 1907 target genes identified in each sequenced isolate using Ridom SeqSphere+ software. Phylogroup analysis was based on in silico PCR using targets identified by Clermont *et al* 2013.

Figure 2. Chromosomal insertion sites of *bla*$_{\text{CTX-M-15}}$.

Chromosomal insertion sites were identified using a combination of miseq and Nanopore Minion single reads sequences. (A) represents the insertion site found in the ST4121 isolate 8 and (B) represents the insertion site found in ST394 isolates 2, 12 and 15. The 11,389bp insertion found in isolates 2, 12 and 15 included *bla*$_{\text{CTX-M-15}}$ and *qnr*S1 and was identical to the first 11,389bp of the 23,174bp insertion in ST4121 isolate 8. The larger insertion in ST4121 isolate 8 included additional *tet*A and *tet*R resistance genes as well as several other genes of plasmid origin. The chromosomal insertion in the ST394 isolates was in a low GC % intergenic region between *ydcs* and *ydcR* genes encoding a putative DNA binding transcriptional regulator and a putative spermidine putrescine transporter subunit, respectively. The insertion in the ST4121 isolate was in a hydrolase gene. Each insertion had the mobile element IS*Ecp1* at the LH terminus and was flanked by target site direct repeats of TATGA for isolate 8 and TTAAA for isolates 2, 12 and 15. This figure appears in colour in the online version of *JAC* and in black and white in the printed version of *JAC*.

References:

1. Abernethy JK, Johnson AP, Guy R, Hinton N, Sheridan EA, Hope RJ. Thirty day all-cause mortality in patients with Escherichia coli bacteraemia in England. Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases. 2015; **21**(3): 251 e1-8.
2. Johnson JR, Nicolas-Chanoine MH, DebRoy C, Castanheira M, Robicsek A, Hansen G, et al. Comparison of Escherichia coli ST131 pulsotypes, by epidemiologic traits, 1967-2009. Emerging infectious diseases. 2012; **18**(4): 598-607.
3. Weissman SJ, Johnson JR, Tchesnokova V, Billig M, Dykhuizen D, Riddell K, et al. High-resolution two-locus clonal typing of extraintestinal pathogenic Escherichia coli. Applied and environmental microbiology. 2012; **78**(5): 1353-60.
4. Clermont O, Christenson JK, Denamur E, Gordon DM. The Clermont Escherichia coli phylo-typing method revisited: improvement of specificity and detection of new phylo-groups. Environ Microbiol Rep. 2013; **5**(1): 58-65.
5. Johnson JR, Tchesnokova V, Johnston B, Clabots C, Roberts PL, Billig M, et al. Abrupt emergence of a single dominant multidrug-resistant strain of Escherichia coli. J Infect Dis. 2013; **207**(6): 919-28.
6. Toleman MA, Bugert, J.J., Nizam, S.A. Extensively drug-resistant New Delhi metallo-β-lactamase–encoding bacteria in the environment, Dhaka, Bangladesh, 2012. . Emerging Infectious Diseases. 2015; **21**(6).
7. Day MJ, Doumith M, Abernethy J, Hope R, Reynolds R, Wain J, et al. Population structure of Escherichia coli causing bacteraemia in the UK and Ireland between 2001 and 2010. The Journal of antimicrobial chemotherapy. 2016; **71**(8): 2139-42.
8. Patzer JA, Walsh TR, Weeks J, Dzierzanowska D, Toleman MA. Emergence and persistence of integron structures harbouring VIM genes in the Children's Memorial Health Institute, Warsaw, Poland, 1998-2006. The Journal of antimicrobial chemotherapy. 2009; **63**(2): 269-73.
9. Sambrook JFaR, D.W. Molecular cloning: a laboratory manual. Third Edition
ed: Cold Spring Harbor Laboratory Press; 2001.
10. Hawser SP, Bouchillon SK, Hoban DJ, Badal RE, Hsueh PR, Paterson DL. Emergence of high levels of extended-spectrum-beta-lactamase-producing gram-negative bacilli in the Asia-Pacific region: data from the Study for Monitoring Antimicrobial Resistance Trends (SMART) program, 2007. Antimicrobial agents and chemotherapy. 2009; **53**(8): 3280-4.
11. Diwan V, Chandran SP, Tamhankar AJ, Stalsby Lundborg C, Macaden R. Identification of extended-spectrum beta-lactamase and quinolone resistance genes in Escherichia coli isolated from hospital wastewater from central India. The Journal of antimicrobial chemotherapy. 2012; **67**(4): 857-9.
12. Kumarasamy KK, Toleman MA, Walsh TR, Bagaria J, Butt F, Balakrishnan R, et al. Emergence of a new antibiotic resistance mechanism in India, Pakistan, and the UK: a molecular, biological, and epidemiological study. Lancet Infect Dis. 2010; **10**(9): 597-602.
13. Perry JD, Naqvi SH, Mirza IA, Alizai SA, Hussain A, Ghirardi S, et al. Prevalence of faecal carriage of Enterobacteriaceae with NDM-1 carbapenemase

at military hospitals in Pakistan, and evaluation of two chromogenic media. The Journal of antimicrobial chemotherapy. 2011.
14. Woodford N, Turton JF, Livermore DM. Multiresistant Gram-negative bacteria: the role of high-risk clones in the dissemination of antibiotic resistance. FEMS Microbiol Rev. 2011; **35**(5): 736-55.
15. Okeke IN, Wallace-Gadsden F, Simons HR, Matthews N, Labar AS, Hwang J, et al. Multi-locus sequence typing of enteroaggregative Escherichia coli isolates from Nigerian children uncovers multiple lineages. PLoS ONE. 2010; **5**(11): e14093.
16. Johnson JR, Porter S, Johnston B, Kuskowski MA, Spurbeck RR, Mobley HL, et al. Host Characteristics and Bacterial Traits Predict Experimental Virulence for Escherichia coli Bloodstream Isolates From Patients With Urosepsis. Open Forum Infect Dis. 2015; **2**(3): ofv083.
17. MacLean RC, San Millan A. Microbial Evolution: Towards Resolving the Plasmid Paradox. Curr Biol. 2015; **25**(17): R764-7.
18. Johnson JR, Johnston BD, Gordon DM. Rapid and Specific Detection of the Escherichia coli Sequence Type 648 Complex within Phylogroup F. Journal of clinical microbiology. 2017; **55**(4): 1116-21.

S2 legend

(Figure 3). This insertion was 11,398 bp and was IS*EcP1* mediated with a typical 5bp target site duplication of host DNA adjacent to the IS*EcP1* left hand terminus (TATGA) and a surrogate IS*EcP1* right hand terminus at the insertion site boundaries. The insertion included *bla*$_{\text{CTX-M-15}}$ and *qnr*S1 resistance genes as well as several partial or complete transposon genes and was similar to *bla*$_{\text{CTX-M-15}}$ containing sections of several plasmids suggesting a plasmid origin. The chromosomal insertion in ST4121 included an identical section of DNA adjacent to IS*EcP1* but was considerably longer including 23,174 bp and included *tet*A and *tet*R genes in addition to *bla*$_{\text{CTX-M-15}}$ and *qnr*S1 genes as well as several other genes of plasmid origin. This insertion was also IS*EcP1* mediated with a typical direct repeat of target DNA at the insertion site (TTAAAA), which was at position 112bp of a hydrolase gene (Figure 3).