Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 May 2020 doi:10.20944/preprints202005.0140.v1

Article

Plasmids shape the diverse accessory resistomes of *Escherichia coli* ST131

Arun G. Decano[1,2], Nghia Tran[3], Hawriya Al-Foori[1], Buthaina Al-Awadi[1], Leigh Campbell[4], Kevin Ellison[1], Louisse P. Mirabueno[1,5], Maddy Nelson[1], Shane Power[1], Genevieve Smith[1], Cian Smyth[1,6], Zoe Vance[7], Caitriona Woods[1], Alexander Rahm[3,8] and Tim Downing[1,*]

[1] School of Biotechnology, Dublin City University, Ireland.

[2] Present address: School of Medicine, University of St. Andrews, UK.

[3] School of Maths, Applied Maths and Statistics, National University of Ireland Galway, Ireland.

[4] Vanderbilt Medical School, Nashville, USA.

[5] Current address: National Institute of Agricultural Botany - East Malling Research, Kent, UK.

[6] Current address: Dept of Biology, Maynooth University, Ireland.

[7] School of Genetics & Microbiology, Trinity College Dublin, Ireland.

[8] Current address: Université de la Polynésie Française, Puna'auia, French Polynesia.

* Correspondence: tim.downing@dcu.ie

Abstract: The human gut microbiome includes beneficial, commensal and pathogenic bacteria that possess antimicrobial resistance (AMR) genes and exchange these predominantly through conjugative plasmids. *Escherichia coli* is a significant component of the gastrointestinal microbiome and is typically non-pathogenic in this niche. In contrast, extra-intestinal pathogenic *E. coli* (ExPEC) including ST131 may occupy other environments like the urinary tract or bloodstream where they express genes enabling AMR and host adhesion like type 1 fimbriae. The extent to which non-pathogenic gut *E. coli* and infectious ST131 share AMR genes and key associated plasmids remains understudied at a genomic level. Here, we examined AMR gene sharing between gut *E. coli* and ST131 to discover an extensive shared preterm infant resistome. In addition, individual ST131 show extensive AMR gene diversity highlighting that analyses restricted to the core genome may be limiting and could miss AMR gene transfer patterns. We show that pEK499-like segments are ancestral to most ST131 Clade C isolates, contrasting with a minority with substantial pEK204-like regions encoding a type IV fimbriae operon. Moreover, ST131 possess extensive diversity at genes encoding type 1, type IV, P and F17-like fimbriae, particular within subclade C2. The type, structure and composition of AMR genes, plasmids and fimbriae varies widely in ST131 and this may mediate pathogenicity and infection outcomes.

Keywords: *Genome*, *fimbrial*, *plasmid*, *ST131*, *Escherichia coli*, *evolution*, *infection*

1. Introduction

Extra-intestinal pathogenic *E. coli* (ExPEC) cause extensive infections outside the gut, from which they can originate. ExPEC have wide-ranging spectra of virulence factors [1-4] and antimicrobial resistance (AMR) [5], especially sequence type (ST) 131 from serotypes O25b:H4 and O16:H5 in phylogroup B2 [6]. ST131 constitutes a substantial minority of ExPEC cases, and possibly a majority of extended-spectrum beta-lactamase (ESBL)-producing ones [7]. ST131's acquisition of key AMR genes encoded on plasmids [4,8] has coincided with its adaptation to new environments [9,10].

ExPEC adherence factors that assist with colonising different host niches are mainly produced by the Chaperone-Usher secretion pathway: type 1, P, F1C/S and AFA fimbriae [11-14]. Within ST131, fluoroquinolone-resistant (FQ-R) Clade C is the main cause of human infection globally [15-16]. The

© 2020 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 May 2020 doi:10.20944/preprints202005.0140.v1

important role of mucosal epithelial binding via the D-mannose-specific type 1 fimbriae encoded by *fim* operon leading to urinary tract or kidney infections has been well established [18-21]. However, the genetic diversity of fimbrial operons beyond *fim* in Clade C has not been examined in large sample collections.

ExPEC infection can be associated with an imbalance in microbiome composition [22] and *E. coli* is the most common initial coloniser of infant intestines [23]. Although most are commensal [24], and others be protective against pathogen invasion via bacteriocin production [25], AMR is prevalent in neonatal units [26] and within-host gene exchange between commensal and pathogenic bacteria may be common [27]. ST131 spreads between mother-infant pairs [28] likely via an oro-faecal transmission route [29], and such colonisations in infants can last for long periods [30]. Consequently, AMR gene screening can inform on pathogen origins, transmission and treatment strategies like collateral sensitivity [31], as well as the effect of antibiotic treatment on microbiome compositions [32].

The ST131 resistome (the set of AMR genes) includes ESBL genes [33-34] that allow third-generation cephalosporin-resistance [35] and are associated with three main cefotaximase (CTX-M) resistance alleles named *bla*$_{\text{CTX-M-14/15/27}}$ [36]. Like most AMR genes, these are encoded on plasmids and are sandwiched by mobile genetic elements (MGEs), and thus are part of the mobile accessory resistome that can be changed by horizontal gene transfer (HGT) [37-39], or lost if no longer beneficial [40-41]. Of nine common bacterial pathogens, *E. coli* had the highest MGE numbers and high levels of transposases and phage-associated MGEs [42]. Diverse MDR bacteria can share MGEs [43-45] and their transfer may make nosocomial outbreaks worse [46-48]. Most plasmids common to ST131 have IS*Ecp1*, IS*903D* and IS*26* elements [49] that encode transpose (*tnpA*) genes flanked by short inverted repeats and are typically adjacent to ESBL genes. Moreover, ESBL genes can be chromosomally transferred and so can form part of the core resistome [50].

Fluctuating antibiotic exposure, host type, anatomical niche and host immunity mean that the resistome within isolates varies [51]. In ST131, plasmid conjugation, plasmid recombination and MGE-mediated rearrangements help shape these resistome dynamics [1,10,52-55]. ST131's plasmids typically are from incompatibility (Inc) groups F, I1, N and A/C [56-58]. Some plasmids in ST131 encode genes for post-segregation killing and stable inheritance to ensure their propagation, but these genes can be lost or may recombine with other plasmids [58-60]. As a result of this mixing and their extensive array of MGEs, plasmids may rearrange extensively even within a clonal radiation [61-62]. Plasmids may also impair cell reproduction due to the energetic cost of their replication and maintenance, and so opportunities for conjugation and recombination could allow for coevolutionary optimisation of gene dosage, coordinated gene expression and recalibrated virulence.

There is extensive evidence of AMR gene conjugation in the gut based on sampled faecal *E. coli* [63]. This operates not only between species [64-66], but also among *E. coli* STs: for instance, a *bla*$_{\text{CTX-M-1}}$-positive IncI1 plasmid was transferred among ST1640, ST2144 and ST6331 within a single patient's gut [67]; a *bla*$_{\text{OXA-48}}$ gene on a IncL/M-type plasmid from *K. pneumoniae* for both ST14 to ST666 [68] and ST648 to ST866 [69]; a 113.6 kb IncF plasmid with a mercury detoxification operon *bla*$_{\text{KPC-2}}$, *bla*$_{\text{OXA-9}}$ and *bla*$_{\text{TEM-1A}}$ genes from ST69 to ST131 [70]; and a range of sulphonamide- and ampicillin-resistance genes (*sul2*, *bla*$_{\text{TEM-1b}}$) were transferred on a pNK29 plasmid within *E. coli* subtypes in the gut [71-72]. Moreover, the frequency of conjugation was ten times higher and more stable in *E. coli* ST131 *bla*$_{\text{CTX-M-15}}$-producing than in *K. pneumoniae* [73].

Recent work has found high accessory genome, plasmid composition and AMR gene diversity in ST131 [50,61], but has yet to examine in detail the mobile resistome and how plasmids' genes interact with chromosomal ones. Here, we examined the preterm infant resistome composition and genomic context across a set of ST131 Clade C isolates in comparison to reference genomes and non-pathogenic samples. We differentiated between the large shared core resistome from the smaller accessory one that varies considerably within closely related isolates. This was caused by distinct events of plasmid turnover and rearrangement, affecting fimbrial adhesion genes as well. We hypothesise that certain plasmids are more compatible with ST131 genomes, and applied topological data analysis (TDA) to understand plasmid-chromosome co-evolution.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 May 2020 doi:10.20944/preprints202005.0140.v1

2. Results

2.1. *A core preterm infant resistome in* E. coli

We examined the resistome and plasmid composition of seven *E. coli* ST131 genome assemblies in the context of related reference genomes and non-pathogenic isolates. The reference resistome in this study was a set of 794 genes on contigs functionally associated with antibiotic resistance in preterm infants (see Data Access) [74] to overcome inconsistencies in existing AMR gene databases (see Methods) and because 79% (n=627 out of 794) of these AMR genes had no clear homology to known ones [74]. This reference resistome came from 401 stool samples longitudinally collected from 84 infants undergoing antibiotic treatment, and was assembled as 2,004 redundant AMR contigs experimentally tested *in vitro* for resistance to 16 antibiotics (Table S1) [74].

The resistome overlap of seven ST131 Clade C (O25b:H4, *fimH30*) genome assemblies originating from human urinary tract infections (UTIs) with subclade C2 reference genome EC958 showed distinct differences in the context of non-pathogenic Clade A reference SE15 (O150:H5, *fimH41*) and two *E. coli* Human Microbiome Project (HMP) assemblies (Table 1). The numbers of unique AMR contigs per assembly ranged from 213 in HMP sample 83972 [75] to 291 in EC958 (Table 1). The non-pathogenic samples tended to have lower AMR gene counts relative to the Clade C ones (Table 1), and more recent ST131 had no difference in AMR gene levels compared to older ones (r^2=0.31, p=0.1), highlighting a need to further explore AMR gene gain and loss.

ID	SRA ID	Clade	*bla*$_{CTX-M}$ allele(s)	Country & year	Unique AMR contigs
8289_1#91	ERR191724	C1	14	Ireland, 2007	268
8289_1#35	ERR191668	C1	14	Ireland, 2008	270
8289_1#3	ERR191636	C2	15	Ireland, 2005	274
8289_1#34	ERR191667	C2	15	Ireland, 2010	274
8289_1#24	ERR191657	C2	14 & 15	Ireland, 2009	283
8289_1#60	ERR191693	C2	15	Ireland, 2010	262
8289_1#27	ERR191660	C2	15	Ireland, 2010	262
SE15	AP009378	A	none	Japan, 2010	251
NCTC13441	ERS530440	C2	15	UK, 2003	280
EC958	HG941718	C2	15	UK, 2005	291
3_2_53FAA	-	ST95	none	Canada, 2007	269
83972	-	ST73	none	Sweden, 1978	213

Table 1. *E. coli* genome assembly collection and metadata. The sample set included seven ST131 short read genome assemblies, three *E. coli* reference genomes, and two HMP assembled contigs. All except the two HMP samples (3_2_53FAA from ST95 and 83972 from ST73) were from ST131. The table shows the sample ID, SRA accession ID, ST131 Clade, *bla*$_{CTX-M}$ allele(s), year of isolation, and numbers of unique AMR contigs. All seven ST131 assemblies, EC958 and NCTC13441 were isolated from pathogenic UTIs. EC958 had 18 AMR contigs on plasmid pEC958A (and none on pEC958B). NCTC13441 had 27 AMR contigs on its plasmid. SE15 was a faecal commensal isolate with had no plasmid-encoded AMR genes. The seven ST131 and two HMP samples were assembled from short Illumina reads, and the three references were assembled from long PacBio reads.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 May 2020 doi:10.20944/preprints202005.0140.v1

2.2. A diverse accessory preterm infant resistome among closely related ST131 isolates

Comparing the AMR contigs' gene product functions identified a core (intrinsic) resistome of 210 AMR contigs shared by diverse samples (Figure 1), highlighting that non-pathogenic isolates like SE15 and the two HMP assemblies from the gut and urinary tract may require these genes to survive. This core resistome differed from the accessory (acquired) resistome of 109 contigs, of which 50 were in pathogenic ST131 only: these corresponded to 18 unique AMR genes (Figure 1). Accessory AMR genes were initially identified with ARIBA v2.11.1 [76] and sequence similarity search comparison against the Comprehensive Antibiotic Resistance Database (CARD) [77] (see Data Access). Results were verified by mapping the read libraries to the reference resistome using GROOT [78] to ensure high read coverage was observed.

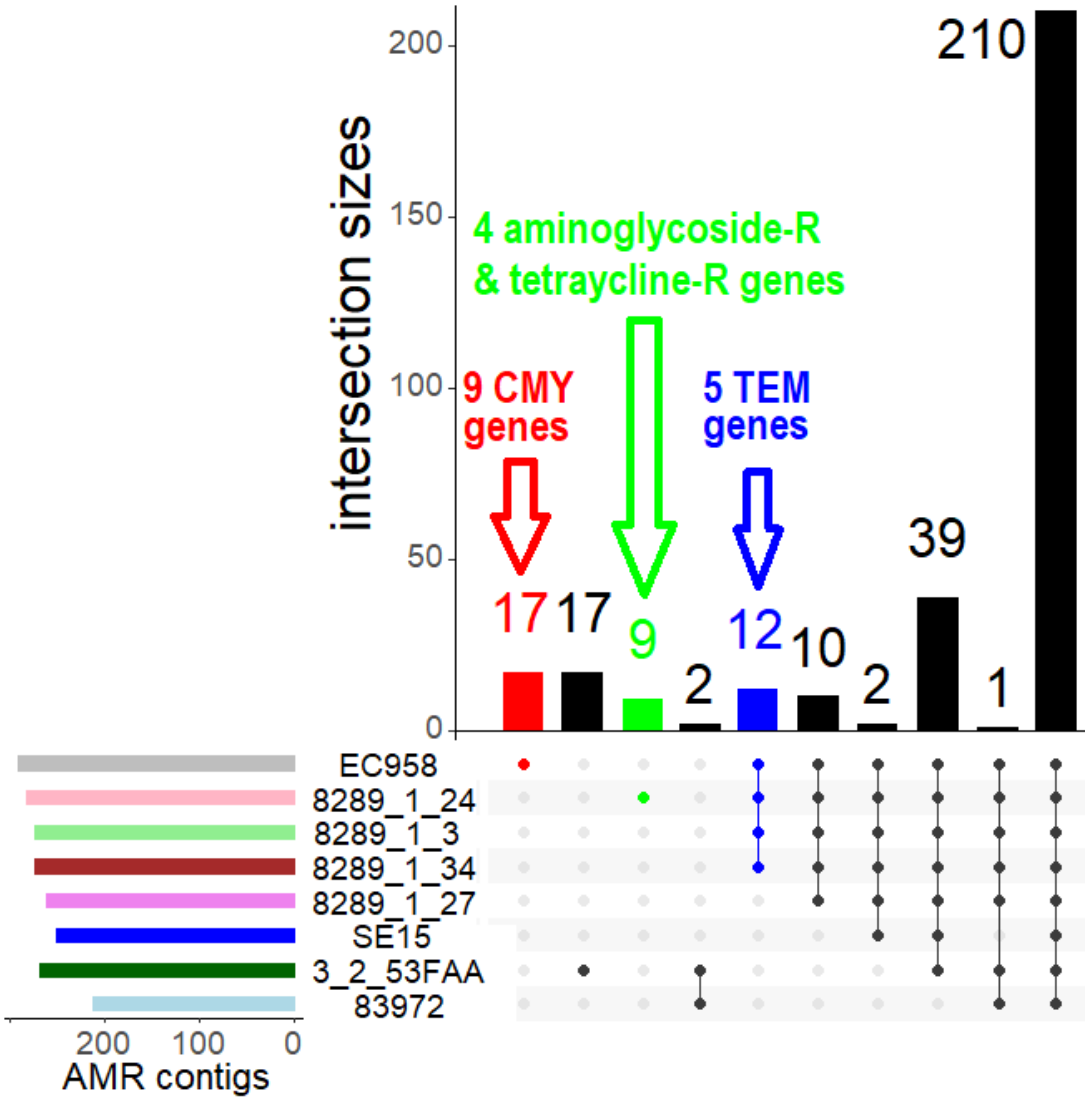

Figure 1. The overlap of preterm infant AMR contigs sharing across four ST131 subclade C2 genome assemblies (8289_1#3, 8289_1#24, 8289_1#27, 8289_1#34), two ST131 reference genomes (EC95 and SE15) and two HMP assemblies (83972 and 3_2_53FAA). Top: The intersection sizes (y-axis) and the numbers of AMR contigs per set showed that most (210) AMR contigs were shared across all isolates. The AMR contigs were redundant and indicated smaller numbers of unique AMR genes, so EC958's 17 AMR contigs (red) corresponded to nine unique *bla*$_{CMY}$ genes (Table S2), and 8289_1#24's nine contigs (green) represented to four unique aminoglycoside and tetracycline resistance genes (Table S3). All Clade C members bar 8289_1#27 had 12 contigs (blue) denoting to five unique *bla*$_{TEM}$ genes (Table S4). Bottom: The numbers of AMR contigs per sample with the corresponding sets shown by black circles. 8289_1#60 (not shown) had the same results as 8289_1#27.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 May 2020 doi:10.20944/preprints202005.0140.v1

Unique to EC958 were nine *bla*CMY alleles resolved from 17 initial AMR contig matches (Figure 1) encoding ESBL AmpC, though cefotaxime-resistance allele *bla*CMY-23 was not among them [79]. These AMR contigs' functional metagenomic screening results showed resistance to penicillins, first- and third-generation cephalosporins (Table S2) [74]. All Subclade C2 isolates bar 8289_1#27 had five *bla*TEM genes associated with cephalosporin resistance (Table S4). The sole *bla*CTX-M-14- and *bla*CTX-M-15-positive isolate (8289_1#24) had unique four genes linked to aminoglycoside and tetracycline resistance (Table S3), a combination that may neutralise the effect of aminoglycoside susceptibility due to tetracycline efflux pump TetA [80-81]. These *bla*TEM, aminoglycoside and tetracycline alleles were associated with penicillin and first-generation cephalosporin resistance [74].

Additional comparisons using the NCTC13441 genome (Table S5) showed that it had five *bla*OXY genes experimentally linked by Gibson et al. (2016) [74] to monobactam and tetracycline resistance in addition to the five *bla*TEM ones (Table S6, Supplementary Results). Ten AMR contigs in all C2 isolates corresponded to a single gene encoding a helix-turn-helix-like transcriptional regulator (Figure 1). The two contigs in all ST131 encoded penicillin-resistance genes, one similar to transcriptional regulator *araC* and the other to *sugE* encoding a protein with resistance to quaternary ammonium salts. The 39 AMR contigs in all assemblies bar 83972 (Figure 1) may enable resistance to amphenicols, penicillins, first-generation cephalosporins and tetracycline, and included 28 copies of *mdfA* encoding a multi-drug efflux pump linked to broad spectrum AMR [82] (Table S7). 3_2_53FAA and 83972 shared one contig region associated with ceftazidime resistance with high similarity to a *ftsI* gene encoding a transpeptidase catalysing peptidoglycan crosslinking [83] that is also called penicillin-binding protein 3 (PBP3) [84-85] and can be inactivated by ESBL binding [86].

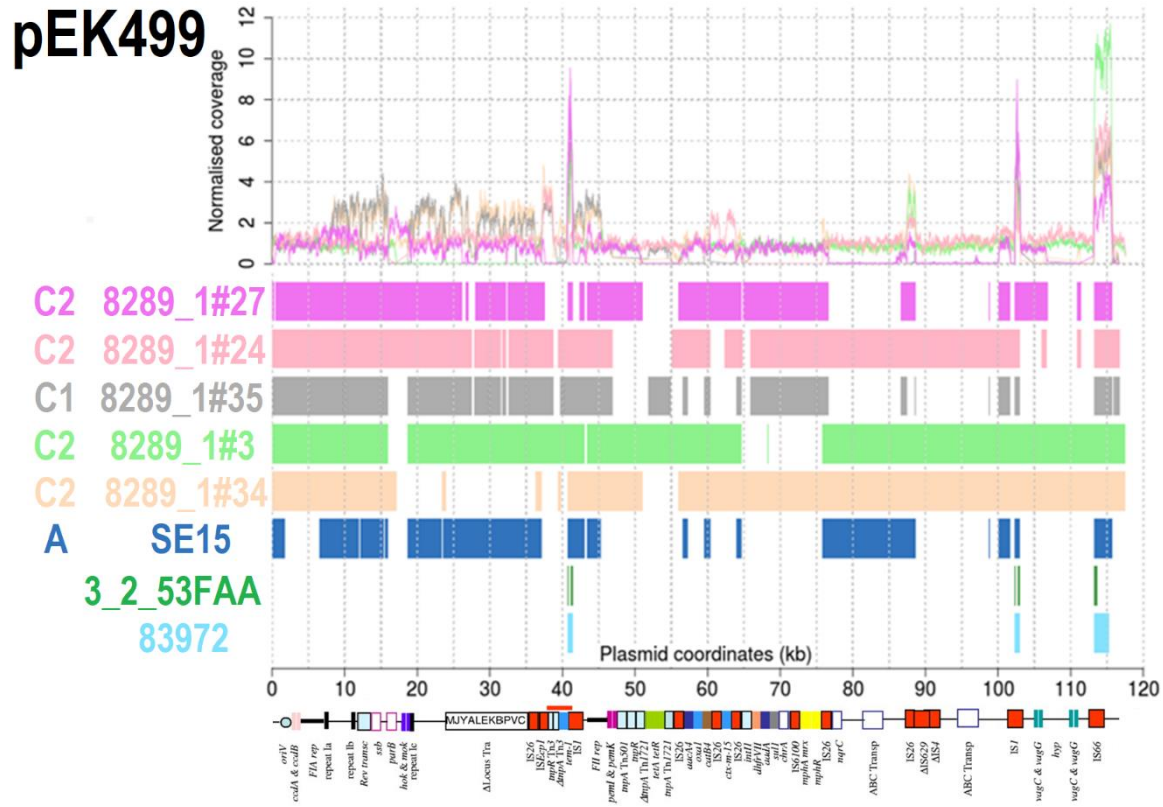

Figure 2. Comparison of pEK499 (117,536 bp) with five ST131 Subclade C1 and C2 isolates, ST131 Clade A reference SE15 (navy) and two HMP assemblies (3_2_53FAA in dark green and 83972 in cyan). Top: Normalised read coverage showed high copy numbers of IS*1* (at 41 and 103 Kb) and IS*66* (at 113 Kb) elements. Bottom: BLAST alignments showed limited similarity for the HMP assemblies, and low matching for SE15 relative to high levels for Clade C assemblies: 8289_1#27 (C2, mauve), 8289_1#24 (C2, pink), 8289_1#35 (C1, grey), 8289_1#3 (C2, light green) and 8289_1#34 (C2, beige). Genes encoding *bla*TEM, *bla*OXA-1 and *bla*CTX-M-15 were at 40, 58 and 63 Kb (respectively). The annotation was modified from [59]. Matches spanning >300 bp are shown.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 May 2020 doi:10.20944/preprints202005.0140.v1

The alignment and mapping results showed that Clade C assemblies had far more regions similar to IncF2/F1A *bla*$_{CTX-M-15}$-positive plasmid pEK499 than SE15 and the HMP samples, but with extensive variation (Figure 2). This plasmid has 185 genes, lacks *traX* for conjugation, is stably inherited (Table S9), and may have been gained early in the emergence of Clade C where it originated as a fusion of type F2 and F1A replicons [88]. This plasmid has ten key AMR genes, nine of which are in a 25 Kb segment demarcated by *bla*$_{TEM}$, *bla*$_{OXA-1}$ and *bla*$_{CTX-M-15}$ at 40, 58 and 63 Kb (respectively). In C2, 8289_1#3 and 8289_1#24 had all three genes, 8289_1#27, 8289_1#60 and 8289_1#34 had *bla*$_{OXA-1}$ and *bla*$_{CTX-M-15}$ but not *bla*$_{TEM}$, whereas the C1 genomes (8289_1#35 and 8289_1#91) were *bla*$_{TEM}$- and *bla*$_{CTX-M-14}$-positive. The conjugation (*tra*) genes at 22-36 Kb were in all ST131 except 8289_1#34.

For a wider context on ST131 dynamics, comparison of the preterm infant resistome with *Klebsiella pneumoniae* showed that the interspecies core resistome had 58 chromosomal genes and that the *K. pneumoniae* core resistome of 308 genes included *bla*$_{SHV}$ (Figure S1, Supplementary Results). Like NCTC13441, nine of the ten genes unique to pathogenic *K. pneumoniae* isolate PMK1 but absent from urinary tract microbiome WGLW1 were *bla*$_{OXY}$ alleles. Repeating the comparison with the more divergent *Staphylococcus aureus* and *lugdunensis* showed no interspecies core resistome (Figure S2) but did find 14 AMR genes on *S. aureus* plasmids (Table S8).

2.3 Extensive plasmid rearrangements in closely related ST131 Clade C genomes

To evaluate the origin and relationships of the preterm infant resistome in ST131 Clade C (Table 1), we examined plasmids common to ST131 (pEK499, pEK516, pEK204, pCA14, pV130) along with SE15, two HMP isolates and CARD. The Clade C read libraries were mapped to these plasmids to infer local gene copy numbers. Plasmids pEK499 (isolated in England), pEK516 (England) and pEK204 (Belfast) were geographically near the other ST131 that were all isolated in Ireland. SE15's pECSF1 (aka pSE15) 122.3 Kb IncF2A/F1B plasmid served as a contrast because it has no AMR genes [87]. 8289_1#35 and 8289_1#91 (both from subclade C1) had identical results, and *vice versa* for 8289_1#60 and 8289_1#27 (both from C2) and so are not shown below.

The IncF2A plasmid pEK516 has ~75% similarity to pEK499 but is shorter (64.6 Kb) [59] and is usually non-conjugative [89]. This plasmid has 103 genes including *bla*$_{TEM}$, *bla*$_{OXA-1}$ and *bla*$_{CTX-M-15}$, and a type I partitioning locus (*parM* and *stbB*, absent in pEK499) for stable inheritance (Table S9). Read mapping to pEK516 showed that Clade C had high similarity to this plasmid unlike SE15 and the HMP samples (Figure S3). Regions at pEK516's *bla*$_{OXA-1}$, *bla*$_{CTX-M-15}$ and *bla*$_{TEM}$ genes at 12, 20 and 24 Kb (respectively) had variable copy numbers.

Contrasting sharply with pEK499 and pEK516, reads mapped to conjugative IncI1 plasmid pEK204 found 72.8 Kb regions of similarity in 8289_1#27 (C2) where the other isolates had <10 Kb of total matching regions (Figure 3). Plasmid pEK204 is similar to IncI1 plasmid R64, has 112 genes, a complete *tra* region (Table S9), and a type I partitioning locus [59]. 8289_1#27 had a complete I1 replicon, *oriT*, *tra*, shufflons subject to site-specific recombinase activity, and a *pil* cluster encoding a type IV fimbriae associated with enhanced cell adherence and biofilm formation in entero-aggregative and Shiga toxin-producing *E. coli* [90]. Regions absent from 8289_1#27 encoded polymerase and UV protection genes (at 13-15 Kb) and the type I partitioning locus (at 18-20 Kb). A 9.3 Kb region on pEK204 contained *bla*$_{TEM-1b}$, along with an inactive transposase Tn*3*-*tnpA*, IS*Ecp1*, *bla*$_{CTX-M}$ and a 5′ orf477-*tnpA*-*tnpR* region. The 14 bp IRL 5′ of IS*Ecp1* and IRR at the distal end of the inverted *orf477* element can mobilise *bla*$_{CTX-M}$, and an additional IRR at *impB* 3′ of the *bla*$_{TEM-1b}$ gene (7.4 Kb further away) can mobilise the 9.3 Kb unit [91], which has been found on IncFIA, IncFIA-FIB, IncN and IncY plasmids and arose on pCOL1b-P9-like plasmid [59]. *ImpB* and *impA* encode an error-prone DNA repair protein (like UmuDC) [92].

Further investigation of the ST131 plasmid-matching regions showed diverse origins with variable plasmid composition within closely related isolates. Read mapping to conjugative plasmid pCA14 showed that all Clade C isolates bar 8289_1#3 had *Mrx* and *mph(A)* genes associated with macrolide resistance (Figure S4). This was also found in comparison with contigs from non-conjugative plasmids pV130a and pV130b from sewage treatment plant water in India [93], where

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 May 2020 doi:10.20944/preprints202005.0140.v1

8289_1#27 and 8289_1#60 (both C2) had similarity spanning the complete length of the pV130a contigs (Figure S5). Inc groups identified using PlasmidFinder [94] showed that all Clade C had IncF1A, but some C2 had lost IncF1B (Table S10). IncF2 and IncX1 were unique to C1, and 8289_1#27 alone had an IncI1 replicon. IncQ1 was in selected C2 samples, as was IncX4 in both C1 and C2.

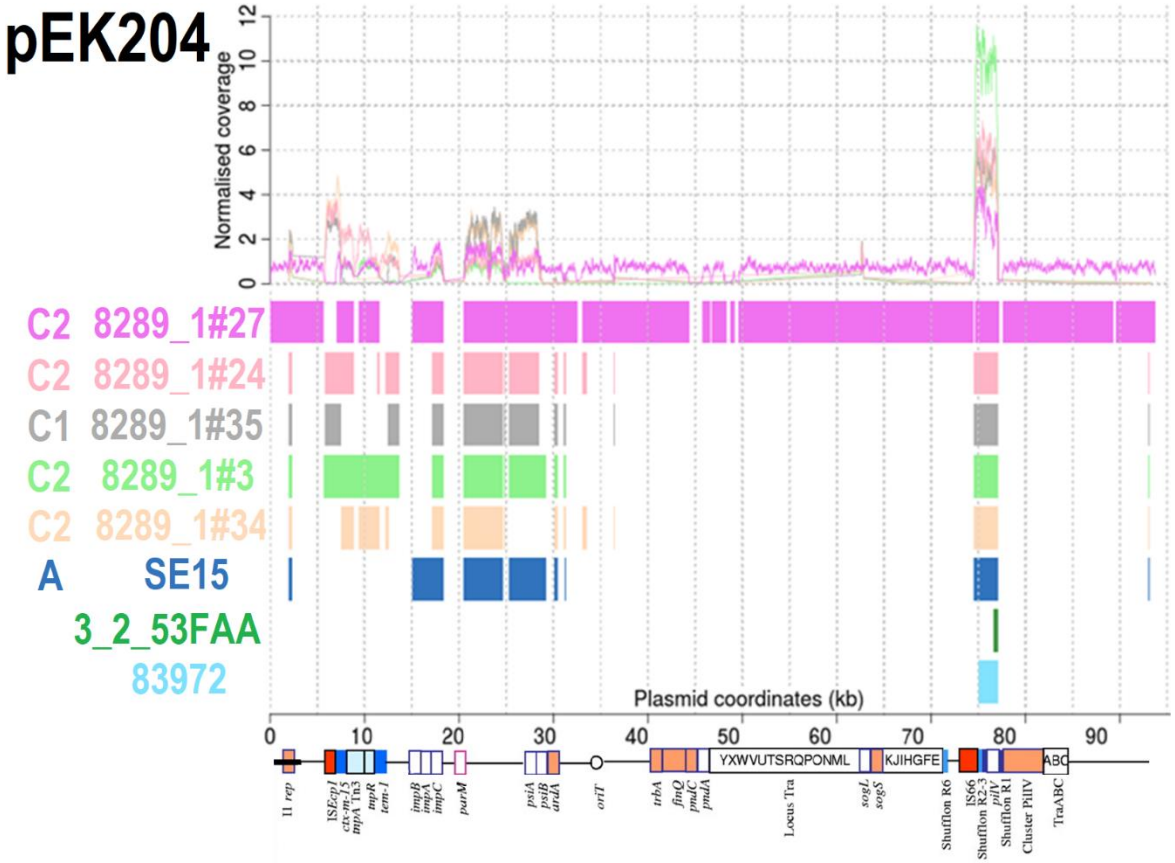

Figure 3. Comparison of pEK204 (93,732 bp) with five ST131 Subclade C1 and C2 isolates, Clade A reference SE15 (navy) and two HMP assemblies (3_2_53FAA in dark green and 83972 in cyan). Top: Normalised read coverage showed high copy numbers at 75-77 Kb encoding an IS*66* and unannotated genes. Bottom: BLAST alignments showed limited similarity for the HMP assemblies, SE15 and Clade C assemblies: 8289_1#24 (C2, pink), 8289_1#35 (C1, grey), 8289_1#3 (C2, light green) and 8289_1#34 (C2, beige). In contrast, 8289_1#27 (C2, mauve) had many regions similar to pEK204. The *bla*$_{\text{CTX-M}}$ gene was at 8 Kb, *bla*$_{\text{TEM}}$ was at 13 Kb, followed by mixed conjugation and segregation genes at 36-70 Kb. The *bla*$_{\text{CTX-M-3}}$ gene typically associated with pEK204 differs from *bla*$_{\text{CTX-M-15}}$ by a R240G substitution, and so the *bla*$_{\text{CTX-M}}$ genes detected for samples here were *bla*$_{\text{CTX-M-15}}$ (not *bla*$_{\text{CTX-M-3}}$). Matches spanning >300 bp are shown. The annotation was modified from [59].

2.4 High rates of pEK499 and pEK204 protein-protein interactions with chromosomal E. coli *proteins*

The extensive diversity in the Clade C accessory resistome raised questions on how their plasmids' gene products interact with chromosomally encoded ones. Protein-protein interaction (PPI) networks can examine plasmid-chromosome coevolution based on gene products' topological proximity [95-96], which should be higher for plasmid and chromosomal proteins that have co-existed for longer periods of time and so interact more often [97]. We used topological data analysis (TDA) to measure the number of non-trivial (indirect) loops where a loop is a chain of at least four PPIs ending at the same protein where it has started. We focus on a type of loops called "non-trivial" (see Methods for details). The number of non-trivial loops was scaled by the number of PPIs in the dataset (Table 2).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 May 2020 doi:10.20944/preprints202005.0140.v1

PPI data for 4,146 *E. coli* protein-coding genes with 105,457 PPIs (all combined scores >400) from the Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) database v10 [98] was processed to get the numbers of non-redundant PPIs and loops per PPI for a plasmid's proteins among themselves, and then between the plasmid's proteins and all 4,146 chromosomal *E. coli* proteins. This included assessing the numbers of PPIs up to a combined score threshold of 900 with a step size of 25. We tested our approach using a published *E. coli* clustering of 60 genes [99] that had a rate of 0.244 non-trivial loops per PPI, which was lower than the rate of 0.518 obtained for all 4,146 chromosomal proteins (Table S11, Figure S6). There were lower rates of non-trivial loops per PPI for pEK499 (0.117) and pEK204 (0.096) compared to the chromosomal rate (0.518), but the plasmids' rates were almost constant across the range of combined score thresholds used (Figure S9). This held even though the number of PPIs per protein was negatively correlated with the combined score (Figure S10), suggesting that the rate of non-trivial loops per PPI was a robust measure of connectivity. Whereas pEK499 and pEK204 interacted with chromosomally encoded proteins, pCA14, pEC958A and pECSF1 did not (Table S12). Plasmid pJIE186-2 did have a small number of PPIs (Table 6) that may be associated with virulence [100].

Plasmid	Number of genes	Number of unique genes	Length (Kb)	Number of unique PPIs		Non-trivial loops per PPI
				Within plasmid	With chromosome	
pEK204	112	87	93.7	21	548	0.096
pEK499	185	87	117.5	8	758	0.117
pEK516	103	55	64.6	8	758	0.117
pJIE186-2	138	78	137.7	2	32	0.015
pCA14	181	93	155.4	5	0	0
pEC958A	142	75	135.6	5	0	0
pECSF1	150	70	122.3	0	0	0

Table 2. The numbers of protein-coding genes per plasmid, unique genes per plasmid, length, and numbers of non-redundant PPIs within the plasmid and with the chromosome, and the rate of indirect interactions (non-trivial loops) per PPI. Plasmid pEK516 had the same PPI results as pEK499.

2.5 ST131 genomes had an ancestral pEK499-like plasmid but some gained a pEK204-like one

The seven ST131 genome assemblies above were among 4,071 [61]. Homology searches of these 4,071 showed 837 (20.6%) had >10 Kb of regions similar to pEK204 (Figure S9), but 3,108 (76.3%) with >10 Kb of regions matching pEK499 (Figure S10). 193 assemblies had >40 Kb of pEK204-like segments, all from Clade C: 17 from C0 (out of 52, 33%), 82 from C1 (out of 1,119, 7%) and 94 from C2 (out of 2,051, 5%) (Figure 4). All 17 from C0 had an I1 replicon, *bla*$_{\text{CTX-M}}$ gene, *bla*$_{\text{TEM}}$ gene (bar one isolate) and at least a partial *tra* region, and 13 had the *pil* operon (76%). The 82 from C1 and 94 from C2 had lower rates of I1 replicon presence (C1 77%, C2 81%), partial *tra* regions (C1 94%, C2 86%) and *pil* clusters (C1 n=60 or 73%, C2 n=71 or 76%) but differed in the rates of gene presence for *bla*$_{\text{CTX-M}}$ (C1 28% vs C2 83%, odds ratio = 12.5, 95% CI 6.1-25.8, p=e-11) and *bla*$_{\text{TEM}}$ (C1 59% vs C2 30%, odds ratio = 3.3, 95% CI 1.8-6.2, p=2.2e-5). This implied *bla*$_{\text{CTX-M}}$ (via *orf477*) was common in C2 and *bla*$_{\text{TEM}}$ (via *impB*) in C1 due to different ancestral transposition of the 9.3 Kb region.

To resolve the origin of pEK204-related *pil* HGT, we searched individually for the 14 *pil* operon genes *pilIJKMNOPQRSTUV* in the 4,071 ST131 assemblies [61]. These genes were conserved as a unit in 376 isolates (9%), including 61 from A (out of 414, 15%), 97 from B (out of 433, 22%), 17 from C0 (out of 52, 33%), 95 from C1 (out of 1,121, 8.5%) and 106 from C2 (out of 2,051, 5.2%) (Table 3) (see Data Access). Tallying this with the pEK204 matching patterns indicated potential for non-pEK204 *pil* gain in Clades A and B, the putative Clade C ancestor may have had a pEK204-like plasmid with no *pil* region, which was sometimes subsequently gained by recombination.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 May 2020 doi:10.20944/preprints202005.0140.v1

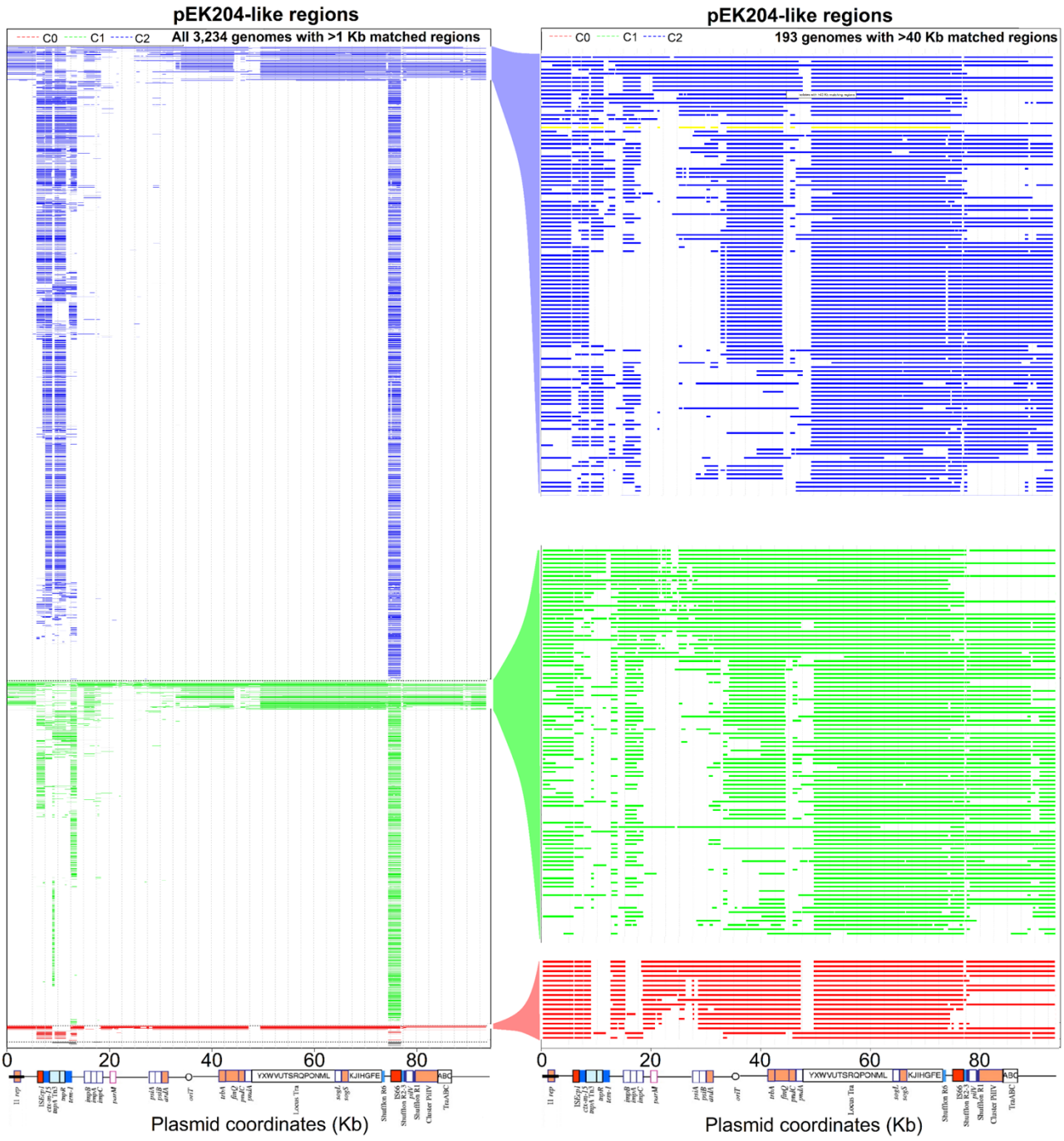

Figure 4. The distributions of pEK204-like regions in ST131 Clade C genome assemblies with >1 Kb (left, n=3,234) and >40 Kb (right, n=193) of matching segments. The regions of similarity were examined in (left) Subclades B0 (n=14 in black), C0 (n=51 in red), C1 (n=1,119 in green) and C2 (n=2,051 in blue), and (right) C0 (n=17 in red), C1 (n=82 in green) and C2 (n=94 in blue). 8289_1#27 (C2) is highlighted in yellow. Similarity across pEK204 (93,732 bp) was based on BLAST alignment. The *bla*$_{\text{CTX-M}}$ gene was at 8 Kb, *bla*$_{\text{TEM}}$ was at 13 Kb, followed by a mix of conjugation and segregation genes at 36-70 Kb. This suggested independent integrations and rearrangements of pEK204-like plasmids across the subclades. The annotation was modified from [59].

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 May 2020 doi:10.20944/preprints202005.0140.v1

Operon			Subclade										
pil	*pap*	*ucl*	A	%	B	%	C0	%	C1	%	C2	%	Totals
-	*papIBA*	-	279	67%	107	25%	3	6%	772	69%	1,192	58%	2,353
-	all	all	21	5%	13	3%	23	44%	82	7%	481	24%	620
all	*papIBA*	-	57	14%	2	<1%			75	7%	66	3%	200
-	all	-	1	<1%	94	22%			8	<1%	14	<1%	117
all	-	-	3	<1%	71	16%			5	<1%	3	<1%	82
all	all	all			16	4%	1	2%	10	<1%	33	2%	60
all	all	-	1	<1%	6	1%	16	31%			4	<1%	27
all	-	all							1	<1%			1
-	*papIBA*	all									1	<1%	1
all	*papIBA*	all											0
-	-	all											0
Partial matches			52	13%	124	29%	9	17%	168	15%	257	13%	610
Totals			414		433		52		1,121		2,051		4,071

Table 3. *Fim*, *pil*, *pap* and *ucl* operon genes detected in 4,071 ST131 genome assembles. The numbers and percentages of isolates per subclade with complete or absent (-) *pil* (*pilK-pilV*) or *pap* or *ucl* (*uclABCD*) operon genes where a subset also had three *pap* genes (*papIBA*). This focused on isolates allocated to the eleven main allelic options, isolates with ssome level of partial operon presence or absence were too sparse to be informative. *UclBCD* but not *uclA* was in an additional 2% (n=77), including 7 from A (2%), 28 from B (6%), none from C0, 24 from C1 (2%) and 18 from C2

Given the *fim* operon's key role in ST131 evolution, we verified that *fim* was present intact in >99% of the 4,071. As expected, the IS*Ec55* insertion at *fimB* [101] that delays *fim* expression in Clade C isolates like EC958 [20] was nearly fixed in C1 (1,117, 99.6%) and C2 (2,041, 99.5%), but was rarer in A (236, 57%), B (93, 21%) and C0 (18, 35%) (including 32 samples with additional *fimB* rearrangements). Isolates with intact *fimB* genes were common in A (178, 43%), B (321, 74%) and C0 (34, 65%).

We extended this to two other important fimbrial operons: *pap* (pyelonephritis-associated pili) and *ucl* (uroepithelial cell adhesin-like). All four *ucl* genes were in 17.5% (n=712) of isolates, including 21 from A (5%), 30 from B (7%), 25 from C0 (48%), 94 from C1 (8%) and notably 542 from C2 (26%) (Table 3). All the *ucl*-positive C2 genomes had a complete *pap* operon (with a single exception). The *pap* operon was in a minority (26%, n=1,049) of isolates, including 28 from A (7%), 163 from B (38%), 49 from C0 (94%), 127 from C1 (11%) and 682 from C2 (33%) (Table 3). Most isolates had *papIBA* and not the genes 3' of them (*papHCDJKEF*) (65%, n=2,661): 359 from A (87%), 118 from B (27%), 3 from C0 (6%), 888 from C1 (79%) and 1,293 from C2 (63%). Moreover, *papI* (234bp) was amplified to two, three of four copies in 22% of these ST131: 26 in Clade A (6%), 51 in B (12%), 45 in C0 (87%), 119 in C1 (11%) and 658 in C2 (32%): more PapI was associated with more P fimbriae expression in *E. coli* CFT073 [102]. Only 45 isolates across clades had an amplified *papB*, and 265 assemblies had a duplicated or triplicated *papA*. These two genes (*papBA*) share a promoter *pBA*, and *papI* has a separate one (*pI*).

This highlighted that most isolates from A, C1 and C2 and a minority of B had no *pil* nor *ucl* operons and lost the *papH-F* segment (Table 3). Clade B differed from A because some from this heterogeneous clade have either *pil* or *pap*, but not both. 83 isolates with no detected *pap* genes all had a *pil* operon, and most (71) were from Clade B. Clades A and B have *pil* independently of pEK204, so Clade C isolates may have gained *pil* on multiple occasions. The Clade C ancestor during the lengthy period prior to the divergence of the rarer C0 subclade from the FQ-R C1/C2 ancestor [50] likely had a complete *pap* operon. This C1/C2 ancestor gained the *fimB* IS*Ec55* insertion and a minority of C1

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 May 2020 doi:10.20944/preprints202005.0140.v1

gained either *pil* (80, 7%) or *ucl* (82, 7%) or both (11, 1%), whereas a sizeable minority of C2 gained *ucl* (482, 24%), rather than *pil* alone (73, 4%) or both together (33, 2%).

Unlike *ucl* gene products that are functionally independent [103], regulatory protein PapB reduces *fim* operon expression by inhibiting FimB and activating FimE, both tyrosine site-specific recombinases that invert *fimS*, including in *E. coli* isolates CFT073 and 536 [104-106]. PapI and PapB regulate *pap* expression [105-107] depending on their concentrations and protein binding at the 416 bp regulatory region between *papI* and *papB*. Thus, retention of *papIBA* in >65% of ST131 could be linked to *fim* transcriptional regulation via *pap*B. Using the STRING PPI network data above for *E. coli* samples 536 and CFT073 that have the *fim*, *pap* and *ucl* operons [108-109] (no data for *pil* gene products was found). There were ten *fim-pap* inter-operon PPIs in both 536 and CFT073, including pilus rod subunit PapA with FimD and FimF (as well as FimC), potentially matching the PPIs of FimA [110]. If only *papIBA* was present, the functional effects (if any) of *papI* and *papA* were uncertain: it is not clear that at the inner cell membrane FimC (replacing PapD) could mediate periplasmic transport of PapA (like FimA) subunits via the chaperone FimI (instead of PapH) to usher FimD (rather than PapC) for pilus rod assembly at the outer cell membrane, extended later by FimF (in lieu of PapK) for the base of the tip. Nonetheless, presence of these genes can inform pilicide and coilicide design, such as anti-PapA antibodies [110].

3. Discussion

AMR genes in commensal or environmental bacteria are a major reservoir for pathogens [111]. This gene flow is driven by HGT and recombination in bacteria [112], including in *E. coli* [113-116], and microbiome species [117-118]. Previous work isolated antibiotic-resistance contigs from preterm infants that were functionally screened for resistance to 16 antibiotics [74]. We showed that this resistome was shared extensively between ExPEC *E. coli* ST131 and microbiome *E. coli* assembled contigs, indicating likely transfer and common ancestry of these genes across commensal and pathogenic bacteria inhabiting the human gut and urinary tract, as expected [119].

Microbiome *E. coli* HMP 83972 (from 2007) does not express functional fimbrial adhesins and is used for therapeutic urinary bladder colonisation in patients in which it protects against super-infection by more virulent strains [120-122]. Although 83972 has lost virulence during its adaptation to commensalism [123], it had key AMR genes like PBP3, which it shared with a microbiome assembly isolated in 2007 (3_2_53FAA). This highlighted the long-term retention of certain AMR genes in asymptomatic specimens, and could provide a contrast for assessing the virulence, biofilm formation, and intestinal colonisation rates that could be lower in ST131 Clade C relative to Clade B [124] but not always [125].

Within ST131, Clade C core genomes are highly conserved and the accessory genomes have extensive differences in AMR gene content [50,52,61]. Here, this was evidenced by NCTC13441's numerous *bla*$_{\text{TEM}}$ and *bla*$_{\text{OXY}}$ genes, contrasting with EC958's abundance of *bla*$_{\text{CMY}}$ ones, with equivalent accessory genome diversity in the seven other Clade C genomes initially examined here. This reinforced the perspective that ST131's accessory resistome is shaped by the environment mediated by plasmids, rather than population structure or geographic factors [126]. Accessory genome screening to develop diagnostics tracking plasmids and MGEs in addition to AMR genes [42,127] could assist with clinical understanding of bacterial infection [128].

In this study, combining the seven Clade C genomes with 4,064 assemblies showed that most (78%) ST131 had >10 Kb of regions similar to plasmid pEK499 and some (9%) also had regions with high similarity to IncI1 plasmid pEK204, suggesting either discrete gains of pEK204-like plasmids with each C subclade or its presence in the Clade C ancestor. IncI1 plasmids are common in *E. coli* clinical infections and are associated with varied ESBL genes [129-131]. Previous work has shown transfer of a *bla*$_{\text{CTX-M-1}}$-positive IncI1 plasmid between *E. coli* STs in the human gut [67]. Given that the mosaic array of backbone plasmid genes may determine fitness effects more so than individual

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 May 2020 doi:10.20944/preprints202005.0140.v1

ESBL genes [132], and evidence of long-term IncF plasmid maintenance in ST131 [62], plasmid pEK204, (pEK516,) and pEK499 proteins' higher interaction rates with chromosomally-encoded proteins relative to other relevant plasmids could represent a co-evolutionary signal. Measuring the indirect connectivity as PPI network loops per interaction showed evidence of this long-term retention that may be relevant for identifying plasmid compatible with *E. coli* chromosomes.

ST131's fitness advantage is tightly correlated with type 1 fimbriae variant *fimH30* and delayed *fim* operon expression due to an insertion at *fimB* [16,133-134]. Here, we found that a minority of ST131 have type IV pilus biosynthesis (*pil*) genes and most of those from Clade C with *pil* also had >40 Kb of regions similar to this pEK204-like plasmid, whereas Clade A and B isolates with *pil* did not have pEK204-like tracts. The *pil* operon may allow different epithelial cell adhesion via a thinner pilus and biofilm formation [90]. Like FimH, UclD is a two-domain tip adhesin that binds intestinal epithelial cells via O-glycans - FimH does this via N-glycans [103]. The F17-like *ucl* operon was more common in subclade C2 at 24%: this operon is associated with uroepithelial cell adhesion, biofilm formation and could be mobilised by its flanking by insertion sequence elements [135].

The *pap* operon encodes a P fimbriae with high specificity for kidney epithelial cell and erythrocyte receptor glycolipids [136]. Moreover, delayed *fim* expression in Clade C isolates [20,101] may be reduced further by PapB: *papIBA* alone from that operon was retained in at least 65% of Clade C isolates here. Delayed *fim* expression is more strongly associated with pyleonephritis than cystitis infections [137], as is having more than one P fimbrial gene cluster [107]. Given that individual *E. coli* express single fimbriae at the cell surface [104,138] and isogenic cell populations express distinct fimbrial types [139], type 1 fimbriae may allow initial bladder colonisation followed by reduced expression in favour of P fimbriae to colonise the kidney [138]. This study's results on plasmid-mediated dynamism across the ST131 resistome and extensive diversity in fimbrial gene composition could inform on potential infection [140] and biofilm formation patterns [20].

4. Materials and Methods

4.1 E. coli *genome isolate collection*

Of the 12 *E. coli* genome assemblies examined (Table 1), three were well-studied ST131 reference genomes: SE15, NCTC13441 and EC958 (Table S5). Seven were ST131 isolated in Ireland during 2005-2010 (Table 2) of which two from subclade C1 were *bla*$_{\text{CTX-M-14}}$-positive, as was 8289_1#24 from Subclade C2, and all five from C2 were *bla*$_{\text{CTX-M-15}}$-positive. All seven were isolated from urine except 8289_1#34, which was a rectal swab. All seven belonged to rST1503 except 8289_1#24 from rST1850 [50]. The remaining two were from the HMP [141] (Table 1): 3_2_53FAA (also called EC3_2_53FAA) was from a colon biopsy in 2007 from a 52-year-old male Canadian with Crohn's disease. 83972 (also called EC83972) was from the urine of a Swedish girl in 1978 with a three-year history of asymptomatic bacteriuria and stable renal function [142], and its core genome has recent common ancestry with the virulent pyelonephritis-causing ExPEC, CFT073 [143] but not avirulent isolates like K-12 MG1655.

Of the three reference genomes, NCTC13441 and EC958 were ST131 subclade C2 isolates from UTIs in the UK, and were assembled from long PacBio reads that had *bla*$_{\text{CTX-M-15}}$-positive plasmids: pEK499 in NCTC13441 and pEC958A in EC958 [9,59,133]. The NCTC13441 genome has 4,983 predicted protein-coding genes and EC958 has 4,982. EC958 has numerous virulence-associated genes that encode adhesins, autotransporter proteins and siderophore receptors, and can cause impairment of uterine contractility in mice [144]. Plasmid pEC958A has 85% similarity with pEK499 but lacks the latter's second *tra* region due to an IS26-mediated *bla*$_{\text{TEM-1}}$ insertion [9,60,133]. The SE15 (O150:H5, *fimH41*) reference genome was examined as a commensal control because this isolate lacks many virulence-associated genes (like α-hemolysin and cytotoxic necrotizing factor), and was a genetic outgroup from ST131 Clade A. It has a 4,717,338 bp chromosome with 4,338 protein-coding genes and a plasmid pSE15 (122 Kb) with 150 protein-coding genes [87].

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 May 2020 doi:10.20944/preprints202005.0140.v1

4.2 Illumina library quality control and read mapping

The paired-end Illumina HiSeq libraries for each ST131 isolate were screened for low quality (Phred score <30) and short (<50 bp) reads using Trimmomatic v0.36 [145] and were corrected using BayesHammer from SPAdes v3.9 [145] (Table 4). These corrected read libraries were mapped to reference sequences with SMALT v7.6 (www.sanger.ac.uk/resources/software/smalt/), and the resulting SAM files were converted to BAM format, sorted and PCR duplicates removed using SAMtools v1.19 [146].

Library	Clade	Number of reads	Read length			ARG-ANNOT	CARD	ResFinder
			Median	Mean	SD			
8289_1#34_1	C2	2,145,896	101	99.3	6.5	9	17	15
8289_1#34_2			100	97.7	8.0			
8289_1#3_1	C2	1,824,030	101	99.4	6.5	17	39	22
8289_1#3_2			100	97.6	8.2			
8289_1#35_1	C1	2,297,624	101	99.4	6.5	44	52	45
8289_1#35_2			100	97.7	8.0			
8289_1#91_1	C1	2,227,251	101	99.3	6.6	44	57	42
8289_1#91_2			100	97.7	8.1			
8289_1#24_1	C2	2,014,250	101	99.3	6.5	56	71	55
8289_1#24_2			100	97.7	8.1			
8289_1#27_1	C2	5,756,848	101	99.4	6.5	6	9	6
8289_1#27_2			100	97.7	8.1			
8289_1#60_1	C2	1,800,070	101	99.3	6.5	14	26	21
8289_1#60_2			100	97.6	8.1			
NCTC13441		2,857,729	43	42.5	1.4			
SE15		418,045	218	192.2	67.3			
EC958		1,514	1,486	1,401.6	549.2			

Table 4. Read library summary statistics for each main ST131 sample and the three main reference genomes. The read distributions differed for NCTC13441, SE15, and EC958 because they were generated using long read approaches. For GROOT usage, paired-end read library files were mapped individually. SD stands for standard deviation. The numbers of AMR genes from ST131 genome assemblies identified in AMR databases ARG-ANNOT, CARD and ResFinder are shown.

4.3 Variable resistome overlaps complicate a clear reference gene set

The reference resistome outlined above was a set of 794 AMR genes on contigs associated with antibiotic resistance in preterm infants assembled as 2,004 redundant AMR contigs experimentally tested *in vitro* for resistance to 16 antibiotics (Table S1) [74]. We compared this set with existing AMR databases to test the robustness of resistome comparisons by aligning the 794 contigs with the CARD [77] and mapping the reads for each of the seven ST131 libraries to the CARD, ARG-ANNOT [147] and ResFinder [148] databases using GROOT [78] where the reads for each isolate were indexed using the median read length (Table 4). This showed considerable variation in the numbers and types of AMR genes determined by each database and extensive AMR gene diversity within Clade C (Table 4). As a negative control, the read library for faecal isolate S250 from ST131 Clade B [149] had no AMR genes in ARG-ANNOT and ResFinder, and five in CARD.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 May 2020 doi:10.20944/preprints202005.0140.v1

4.4 Homology-based resistome screening and comparison

Contig annotation and protein domain recognition were implemented using the Pfam v27.0 and ProSite databases using InterProScan v5.22-61 [150]. The protein homolog dataset in CARD (2,239 genes) [77] was aligned with the isolates' genomes to annotate the resistomes using BLAST v2.2.31 where matches with a bit score >500 and >99% homology were considered valid. Alignments were visualised using R's VennDiagram v1.6.1, Seqinr v3.4-5, UpSetR v1.4.0 and WriteXLS v5.0.0 packages. Detected genes and elements were verified and visualised with Artemis and the Artemis Comparison Tool (ACT) [151].

4.5 Assessment of plasmids prevalent in ST131

Sequence and annotation files for pEK499 (NC_013122.1, EU935739), pEK516 (NC_013121.1, EU935738), pEK204 (NC_013120.1, EU935740) and all pV130 contigs (LC056314.1 to LC056328.1 [93] were downloaded. The latter were aligned to one another to examine their uniqueness with Clustal Omega v1.2.1 [152]. Replicon typing was completed using PlasmidFinder [56] and each plasmid was compared to the CARD as above. Each *E. coli* sequence library was mapped as above to each plasmid to verify local AMR gene and MGE genetic structure and determine copy number levels, which were visualised with Artemis [151] and R v3.4.2's reshape2 v1.4.3, ggridges v0.5.1, ggplot2 v2.3.2.1, readr v1.3.1, dplyr v0.8.3 and ape v5.3 packages. Plasmid and gene sequence similarity was calculated using the Sequence Identity and Similarity (SIAS) tool (http://imed.med.ucm.es/Tools/sias.html). Homology searches for contigs with high similarity to pEK204 and pEK499 in 4,071 assembled ST131 genomes [60] examined matches with >98% sequence identity and a length of >300 bp.

4.6 Protein-protein interaction network construction and topological data analysis

TDA has been used to investigate complex and high-dimensional biological datasets, like breast cancer genomes [153], and here we applied it to PPI networks with a focus on direct and indirect connectivity across different network topologies by quantifying rates of absent PPIs (non-trivial loops). The numbers of non-trivial loops per PPI was used as a metric for indirect connectivity because it was consistent across parameters and was independent of network size. We assessed pCA14, pEK204, pEK499, pEK516 and pEC958A given the immediate relevance to ST131 genome composition, along with two plasmids with no known AMR genes as negative controls: pECSF1 (aka pSE15), and non-conjugative plasmid pJIE186-2 that has numerous virulence factor genes that are chromosomal in some *E. coli* [100]. We examined unique genes with PPI network information only: here, results for pEK516 were identical to those for pEK499.

We extracted *E. coli* K12 MG1655 PPI data from the STRING database v10 [98] with R v3.5.2 packages BiocManager v1.30.4, dplyr v0.8.0.1, genbankr v1.10.0, rentrez v1.2.1, STRINGdb v1.22.0, tidyverse v1.2.1 and VennDiagram v1.6.20. STRING's combined scores were used because they integrate multiple types of evidence while controlling for random PPIs [98]. Using a score threshold of 400+ to avoid spurious PPIs, the K12 dataset had 4,146 distinct protein-coding genes, of which 4,121 had interaction information, resulting in 105,457 PPIs (14,555 PPIs were present for a score threshold of 900). For each plasmid or set of genes, we obtained the unique genes and for these determined the number of pairwise PPIs within that set and with the chromosome. We tested our TDA-based approach using the results of relevant previous work [99]. For *E. coli* 536, FooB was used as an equivalent for PapB, and likewise for protein F7-2 that matched virulence factor PapA in CFT073, and FooG for PapG.

We constructed the Vietoris-Rips complex [154] where the proteins were the vertices with PPIs as edges, so that proteins A and B with an PPI would be joined by a single edge (two proteins in one dimension, 1-D). Similarly, proteins A, B and C that are joined by three PPIs would get their PPI triangle filled in with a solid 2-D triangle. For four proteins A, B, C and D joined by all six possible pairwise PPIs, their four filled 2-D triangles A/B/C, B/C/D, A/C/D and A/B/D constitute a tetrahedral surface

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 May 2020 doi:10.20944/preprints202005.0140.v1

which gets filled in with a solid 3-D tetrahedron. This can be extended to *m+1* pairwise connected proteins, which get a solid *m-D* polytope filled into their skeleton of edges, triangles, tetrahedra, and so on. A loop connects *m* proteins, where *m* is 4, 5, 6 or more, by precisely *m* PPIs, such that each protein is involved in precisely two of these PPIs. For instance, four proteins A, B, C, D can be joined to a loop by the four PPIs A-B, B-C, C-D, D-A, where A-B is a PPI between the proteins A and B (and so on). Any loop which can be filled with triangles, along existing PPIs, is called trivial. So if there is an interaction between A and C, the loop is filled by the two triangles AB-BC-CA and CD-DA-CA and hence trivial. If there are no interactions between A and C, and none between B and D, then the loop is non-trivial.

Due to the large K12 dataset size (4,146 proteins with 105,457 interactions), instead of using general purpose implementations of the Vietoris-Rips complex (such as SAGE), specialised software [155] to optimise the Betti number computations was used. This software used the sparsity of the boundary matrices to process the rank computations efficiently in LinBox [156], with a hard-coded dimensional truncation of the Vietoris-Rips complex to avoid the large number of high-dimension simplices that could be obtained for the full dataset. For each analysis across STRING combined scores 400 to 900 with a step of 25, we computed the first Betti number of the Vietoris-Rips complex that counted the numbers of non-trivial loops (with missing PPIs inside), which was adjusted for the numbers of PPIs above the score threshold (ie, loops per PPI). The number of PPIs for the complete K12 chromosomal dataset was negatively correlated with the combined score threshold (r^2=0.964), as was the loops per PPI (r^2=0.905), consequently the chromosomal loops per PPI rate was used as a baseline metric across different score thresholds.

4.7 Operon gene homology search approach

Homology searches for each *fim* operon gene were implemented using the NCTC13441 annotation coordinates to extract the corresponding sequence with SAMtools v1.9 and align each one to the 4,071 ST131 genome assemblies [60] with BLAST v2.2.31 as above, processed with R packages tidyverse v1.2.1 and dplyr v0.8.3. For all operons, minor individual gene partial matches or losses were not examined due to the large number of samples and lack of consistent patterns for rare combinations. Results on amplified genes were restricted to those with prevalance >1%. The CDS of *fimB* is normally 600 bp, and with the IS*Ec55* insertion the homologous region typically spanned 2,493 bp. NCTC13441 has only three *pap* genes (*papIBA*) and lacks any *ucl* or *pil* genes, so the *pil* genes were extracted from pEK204. *PilI* and *pilJ* encoding IncI1 conjugal transfer proteins were mainly absent with no clear association with the other *pil* genes, and so were not examined in detail here. *PilV* encoding a pilus tip adhesin was at least duplicated in 292 assemblies, as expected for a locus rearranged to change pilus binding specificity [157]. The 5 Kb *ucl* operon has four genes, a major subunit (*uclA*), chaperone (*uclB*), usher (*uclC*) and adhesin (*uclD*), whose reference sequences were determined from *E. coli* 83972 (CP001671). *E. coli* UTI89's chromosome (CP000243) and 143-gene 114 Kb plasmid (CP000244) [112] were used to obtain the *pap* operon: this has regulatory protein gene *papI*, structural protein genes *papA/B* and seven structural subunit genes (*papHCDJKEF*).

Supplementary Materials: The following are available online at www.mdpi.com/xxx/s1:

Table S1 - The 16 antibiotics used to identify 794 AMR genes on contigs originally from preterm infants.
Table S2 - The nine AMR genes unique to EC958 were Ambler class C β-lactamases, that are typically associated with cephalosporin resistance.
Table S3 - The AMR genes unique to 8289_1#24 were linked to aminoglycoside & tetracycline resistance.
Table S4 - The five AMR genes unique to all ST131 Clade C genomes except 8289_1#27 encoded *bla*$_{TEM}$ genes (an Ambler class A β-lactamase).
Text S1 - The preterm infant resistome in ST131 genomes including NCTC13441.
Table S5 - The three *E. coli* reference genomes and the contig assembly of 3_2_53FAA from the Human Microbiome Project (HMP).
Table S6 - The five additional AMR genes unique to NCTC13441.
Table S7 - Seven AMR genes in ST131 and microbiome 3_2_53FAA from 39 AMR contigs.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 May 2020 doi:10.20944/preprints202005.0140.v1

Text S2 - A small shared infant resistome between ST131 and *Klebsiella pneumoniae*
Figure S1 - Chromosomal AMR genes in *K. pneumoniae* reference isolate PMK1 versus a *K. pneumoniae* microbiome (WGLW1) contig set from the urogenital tract, relative to pathogenic *E. coli* ST131 EC958 and commensal *E. coli* SE15.
Text S3 - No shared infant resistome between ST131 and *Staphylococcus lugdunensis*
Figure S2 - Out of 794 preterm infant gut AMR genes, there were no overlapping chromosomal genes in a *S. lugdenensis* reference HKU versus a *S. lugdenensis* microbiome (M23590) contig set, relative to pathogenic *E. coli* ST131.
Table S8 - Ten *S. aureus* reference genome LAC USA300 plasmids had similarity to 13 AMR contigs corresponding to six AMR genes.
Table S9 - Key known AMR, plasmid persistence and conjugation genes for pEK204, pEK499, pEK516, pCA14, pV130a and pV130b.
Figure S3 - Comparison of pEK516 (64,471 bp) with ST131 and two HMP assemblies.
Figure S4 - Read mapping distributions for seven ST131 to pV130 contigs.
Figure S5 - Comparison of plasmid pCA14 (155 Kb) with seven ST131 C1/C2 isolates.
Table S10 - PlasmidFinder Inc group alignments against pV130 and seven ST131 isolates.
Table S11 - The PPIN characteristics of the protein-coding genes from Miryala et al (2018).
Table S12 - The seven plasmids' accessions, numbers of genes with interactions given a combined score of 400+, Inc groups, conjugative ability and AMR genes.
Figure S6 - The ratio of non-trivial loops per interaction versus the STRING combined score for all 60 genes from Miryala et al (2018) Table 1.
Figure S7 - The ratio of non-trivial loops per interaction versus the STRING combined score for all chromosomal genes, all genes on pEK499 and all those on pEK204.
Figure S8 - The ratio of interactions per protein for genes encoded on the chromosome or the pEK499 and pEK204 plasmids across the STRING combined score.
Figure S9 - The distributions of pEK204-like and pEK499-like regions in ST131 assemblies.
Figure S10 - The distributions of pEK204-like regions in ST131 Clade C genome assemblies.

Author Contributions: Conceptualization, A.D., A.R. and T.D.; Bacterial Genome Analysis, A.D., H.A., B.A., L.C., K.E., L.M., M.N., S.P., G.S., C.S., Z.V., C.W.; Topological Data Analysis, N.T., A.R. and T.D.; Writing – Original Draft Preparation, A.D., N.T., A.R. and T.D.; Writing – Review & Editing, A.D. and T.D.; Funding Acquisition, A.D., T.D.

Funding: This work was funded by a DCU O'Hare Ph.D. fellowship and a DCU Enhancing Performance grant.

Conflicts of Interest: The authors declare no conflict of interest.

Data Access:

The following files are available on the FigShare project "Plasmids_ST131_resistome_2020":

1. The set of 794 AMR genes derived from Gibson et al. (2016) are available (with their protein sequence translation) at FigShare at doi: 10.6084/m9.figshare.11961402.

2. The AMR gene profiles per sample determined by their BLAST sequence similarity results against CARD are available at FigShare at doi: 10.6084/m9.figshare.11961612. This dataset includes the PlasmidFinder results. It also includes other AMR database comparisons (ARG-ANNOT, ResFinder, MegaRes, VFDB and VirulenceFinder).

3. The BLAST sequence similarity results for the *fim*, *pil*, *pap* and *ucl* operons' genes versus 4,071 *E. coli* ST131 assemblies from Decano & Downing (2019) are available at FigShare at doi: 10.6084/m9.figshare.11961711.

4. The genome sequences and annotation files for reference genomes NCTC13441, EC958 and SE15, along with the assembled contigs for 83972 and 3_2_53FAA are available at FigShare at doi: 10.6084/m9.figshare.11961813.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 May 2020 doi:10.20944/preprints202005.0140.v1

5. The 4,071 *E. coli* ST131 genome assemblies from Decano & Downing (2019) are available at FigShare at doi: 10.6084/m9.figshare.11962278 (the first 1,680 assemblies) and at doi: 10.6084/m9.figshare.11962557 (the second 2,391 assemblies).

References

1. Ben Zakour, N.L.; et al. Sequential acquisition of virulence and fluoroquinolone resistance has shaped the evolution of *Escherichia coli* ST131. *mBio* **2016** *7*, e00347, https://doi.org/10.1128/mBio.00347-16.

2. Calhau, V.; Ribeiro, G.; Mendonça, N; Da Silva, G.J. Prevalent combination of virulence and plasmidic-encoded resistance in ST 131 *Escherichia coli* strains. *Virulence* **2013** *4(8)*, 726–9, https://doi.org/10.4161/viru.26552.

3. Van der Bij, A.K.; Peirano, G.; Pitondo-Silva, A; Pitout, J.D. The presence of genes encoding for different virulence factors in clonally related *Escherichia coli* that produce CTX-Ms. *Diagn Microbiol Infect Dis* **2012** *72*, 297–302, https://doi.org/10.1016/j.diagmicrobio.2011.12.011.

4. Goswami C.; Fox S.; Holden M.; Connor M.; Leanord A.; Evans TJ. Genetic analysis of invasive *Escherichia coli* in Scotland reveals determinants of healthcare-associated versus community-acquired infections. *Microb Genom.* **2018** *4(6)*, doi: 10.1099/mgen.0.000190

5. Poolman, J.T.; Wacker, M. Extraintestinal Pathogenic *Escherichia coli*, a Common Human Pathogen: Challenges for Vaccine Development and Progress in the Field. *J Infect Dis.* **2016** *213(1)*, 6–13, https://doi.org/10.1093/infdis/jiv429.

6. Livermore D. The 2018 Garrod Lecture: Preparing for the Black Swans of resistance. *Journal of Antimicrobial Chemotherapy* **2018** *73*, 2907–2915, https://doi.org/10.1093/jac/dky265

7. Banerjee, R.; Johnson, J.R. A new clone sweeps clean: the enigmatic emergence of *Escherichia coli* sequence type 131. Antimicrob Agents Chemother **2014** 58, 4997–5004, https://doi.org/10.1128/AAC.02824-14.

8. Pitout J.D.D.; DeVinney R. *Escherichia coli* ST131: a multidrug-resistant clone primed for global domination. F1000 Research **2017** 6(F1000 Faculty Review), 195.

9. Forde, B.M.; et al. The complete genome sequence of *Escherichia coli* EC958: A high quality reference sequence for the globally disseminated multidrug resistant *E. coli* O25b:H4-ST131 clone. Plos One **2014** *9(8)*, pp.e104400.

10. McNally, A.; et al. Combined analysis of variation in core, accessory and regulatory genome regions provides a super-resolution view into the evolution of bacterial populations. PLoS Genet **2016** *12*, 1006280.

11. Kuehn M.J.; Heuser J.; Normark S.; Hultgren S.J. P pili in uropathogenic *E. coli* are composite fibres with distinct fibrillar adhesive tips. Nature **1992** *356(6366)*, 252-5

12. Connell I.; Agace W.; Klemm P.; Schembri M.; Mărild S.; Svanborg C. Type 1 fimbrial expression enhances *Escherichia coli* virulence for the urinary tract. *Proc Natl Acad Sci U S A* **1996** *93(18)*, 9827-321996.

13. Martinez J.J.; Mulvey M.A.; Schilling J.D.; Pinkner J.S.; Hultgren S.J. Type 1 pilus-mediated bacterial invasion of bladder epithelial cells. *EMBO J.* **2000** *19(12)*, 2803-122000

14. Selvarangan R, Goluszko P, Singhal J, Carnoy C, Moseley S, Hudson B, Nowicki S, Nowicki B. Interaction of Dr adhesin with collagen type IV is a critical step in *Escherichia coli* renal persistence. *Infect Immun.* **2004** *72(8)*, 4827-352004.

15. Price, L.B.; et al. The epidemic of extended-spectrum-beta-lactamase-producing *Escherichia coli* ST131 is driven by a single highly pathogenic subclone, H30-Rx. mBio **2013** *4*, e00377–13, https://doi.org/10.1128/mBio.00377-13.

16. Petty N.K.; et al. Global dissemination of a multidrug resistant Escherichia coli clone. *Proc Natl Acad Sci U S A.* **2014** *111(15)*, 5694-9. doi: 10.1073/pnas.1322678111

17. Bloch C.A.; Stocker B.A.; Orndorff P.E. A key role for type 1 pili in enterobacterial communicability. *Mol. Microbiol.* **1992** *6*, 697–701.

18. Bloch C.A.; Orndorff P.E. Impaired colonization by and full invasiveness of *Escherichia coli* K1 bearing a site-directed mutation in the type1 pilin gene. Infect. Immun. **1990** *58*, 275–278

19. Totsika M.; et al. A FimH inhibitor prevents acute bladder infection and treats chronic cystitis caused by multidrug-resistant uropathogenic *Escherichia coli* ST131. *J Infect Dis* **2013** *208*, 921–8.

20. Sarkar S.; Vagenas D.; Schembri M.A.; Totsika M. Biofilm formation by multidrug resistant *Escherichia coli* ST131 is dependent on type 1 fimbriae and assay conditions. *Pathog Dis* **2016** *74* doi: 10.1093/femspd/ftw013.

21. Bahrani-Mougeot F.K.; Buckles E.L.; Lockatell C.V.; Hebel J.R.; Johnson D.E.; Tang C.M.; Donnenberg M.S. Type 1 fimbriae and extracellular polysaccharides are preeminent uropathogenic *Escherichia coli* virulence determinants in the murine urinary tract. Mol. Microbiol. **2002** *45*, 1079 –1093.

22. Bäumler A.J.; Sperandio V. Interactions between the microbiota and pathogenic bacteria in the gut. Nature **2016** *535(7610)*, 85-93. doi: 10.1038/nature18849

23. Bittinger K.; et al. Bacterial colonization reprograms the neonatal gut metabolome. *Nat Microbiol* **2020** doi: 10.1038/s41564-020-0694-0

24. Conway T.; Cohen P.S. Commensal and Pathogenic *Escherichia coli* Metabolism in the Gut. *Microbiol Spectr.* **2015** *3(3)* doi: 10.1128/microbiolspec.MBP-0006-2014

25. Hudault S.; Guignot J.; Servin A.L. *Escherichia coli* strains colonising the gastrointestinal tract protect germfree mice against Salmonella typhimurium infection. *Gut* **2001** *49(1)*, 47-55

26. Li G.; et al. Towards understanding global patterns of antimicrobial use and resistance in neonatal sepsis: insights from the NeoAMR network. *Arch Dis Child.* **2020** *105(1)*, 26-31. doi: 10.1136/archdischild-2019-316816

27. Groussin M.; et al. Industrialization is associated with elevated rates of horizontal gene transfer in the human microbiome. *BioRxiv* **2020** doi: http://dx.doi.org/10.1101/2020.01.28.922104

28. Gurnee E.A.; et al. Gut Colonization of Healthy Children and Their Mothers With Pathogenic Ciprofloxacin-Resistant *Escherichia coli*. J Infect Dis. 2015 212(12), 1862-8. doi: 10.1093/infdis/jiv278

29. Yamamoto S.; Tsukamoto T.; Terai A.; Kurazono H.; Takeda Y.; Yoshida O. Genetic evidence supporting the fecal-perineal-urethral hypothesis in cystitis caused by *Escherichia coli. J Urol* **1997** *157*,1127–9.

30. Rodríguez-Revuelta M.J.; López-Cerero L.; Serrano L.; Luna-Lagares S.; Pascual A.; Rodríguez-Baño J. Duration of Colonization by Extended-Spectrum β-Lactamase-Producing Enterobacteriaceae in Healthy Newborns and Associated Risk Factors: A Prospective Cohort Study. *Open Forum Infect Dis.* **2018** *5(12)*, ofy312 doi: 10.1093/ofid/ofy312

31. Roemhild R.; Linkevicius M.; Andersson D.I. Molecular mechanisms of collateral sensitivity to the antibiotic nitrofurantoin. *PLoS Biol.* **2020** *18(1)*, e3000612. doi: 10.1371/journal.pbio.3000612

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 May 2020 doi:10.20944/preprints202005.0140.v1

32. Russell J.T.; et al. Antibiotics may influence gut microbiome signaling to the brain in preterm neonates. *BioRxiv* **2020** doi: https://doi.org/10.1101/2020.04.20.052142

33. ECDC, European Centre for Disease Prevention and Control. European Centre for Disease Prevention and Control. Antimicrobial resistance surveillance in Europe 2015. Annual Report of the European Antimicrobial Resistance Surveillance Network (EARS-Net). **2017** Stockholm: ECDC.

34. Findlay J.; Gould V.C.; North P.; Bowker K.E.; Williams M.O.; MacGowan A.P.; Avison M.B. Characterization of cefotaxime-resistant urinary *Escherichia coli* from primary care in South-West England 2017-18. *J Antimicrob Chemother.* **2020** 75(1), 65-71. doi: 10.1093/jac/dkz397

35. Mathers A.J.; Peirano G.; Pitout J.D. The role of epidemic resistance plasmids and international high-risk clones in the spread of multidrug-resistant *Enterobacteriaceae*. *Clin Microbiol Rev*. **2015** *28(3)*, 565–91, https://doi.org/10.1128/CMR.00116-14

36. Pitout J.D.; Laupland K.B. Extended-spectrum beta-lactamase-producing Enterobacteriaceae: an emerging public-health concern. *Lancet Infect Dis.* **2008** *8(3)*, 159-66. doi: 10.1016/S1473-3099(08)70041-0

37. Stokes H.W.; Gillings M.R. Gene flow, mobile genetic elements and the recruitment of antibiotic resistance genes into Gram-negative pathogens. *FEMS Microbiol Rev* **2011** *35*, 790-819, doi:10.1111/j.1574-6976.2011.00273.x

38. Von Wintersdorff C.J.; et al. Dissemination of Antimicrobial Resistance in Microbial Ecosystems through Horizontal Gene Transfer. *Front Microbiol* **2016** *7*, 173, doi:10.3389/fmicb.2016.00173.

39. Goldstone R.J.; Smith D.G.E. A population genomics approach to exploiting the accessory 'resistome' of *Escherichia coli*. *Microb Genom.* **2017** 3(4), e000108. doi: 10.1099/mgen.0.000108

40. Baraniak A.; Fiett J.; Hryniewicz W.; Nordmann P.; Gniadkowski M. Ceftazidime-hydrolysing CTX-M-15 extended-spectrum beta-lactamase (ESBL) in Poland. J Antimicrob Chemother. **2002** *50(3)*, 393-6

41. Pehrsson E.C.; Tsukayama P.; Patel S.; Mejía-Bautista M.; Sosa-Soto G.; Navarrete K.M.; Calderon M.; Cabrera L.; Hoyos-Arango W.; Bertoli M.T.; Berg D.E.; Gilman R.H.; Dantas G. Interconnected microbiomes and resistomes in low-income human habitats. *Nature.* **2016** *533(7602)*, 212-6. doi: 10.1038/nature17672

42. Durrant M.G.; Li M.M.; Siranosian B.A.; Montgomery S.B.; Bhatt A.S. A Bioinformatic Analysis of Integrative Mobile Genetic Elements Highlights Their Role in Bacterial Adaptation. *Cell Host Microbe.* **2020** *27(1)*, 140-153.e9. doi: 10.1016/j.chom.2019.10.022

43. Cerqueira G.C.; et al. Multi-institute analysis of carbapenem resistance reveals remarkable diversity, unexplained mechanisms, and limited clonal outbreaks. *Proc Natl Acad Sci U S A.* **2017** *114(5)*, 1135-1140. doi: 10.1073/pnas.1616248114

44. Hazen T.H.; et al. Diversity among bla(KPC)-containing plasmids in *Escherichia coli* and other bacterial species isolated from the same patients. *Sci Rep.* **2018** *8(1)*, 10291. doi: 10.1038/s41598-018-28085-7

45. Kwong J.C.; et al. Translating genomics into practice for real-time surveillance and response to carbapenemase-producing *Enterobacteriaceae*: evidence from a complex multi-institutional KPC outbreak. *PeerJ* **2018** *6*, e4210. doi: 10.7717/peerj.4210

46. Bosch T.; et al. Outbreak of NDM-1-Producing *Klebsiella pneumoniae* in a Dutch Hospital, with Interspecies Transfer of the Resistance Plasmid and Unexpected Occurrence in Unrelated Health Care Centers. *J Clin Microbiol.* **2017** *55(8)*, 2380-2390. doi: 10.1128/JCM.00535-17

47. Jamrozy D.; et al. Evolution of mobile genetic element composition in an epidemic methicillin-resistant *Staphylococcus aureus*: temporal changes correlated with frequent loss and gain events. *BMC Genomics* **2017** *18(1)*, 684. doi: 10.1186/s12864-017-4065-z

48. Martin J.; et al. Covert dissemination of carbapenemase-producing *Klebsiella pneumoniae* (KPC) in a successfully controlled outbreak: long- and short-read whole-genome sequencing demonstrate multiple genetic modes of transmission. *J Antimicrob Chemother.* **2017** *72(11)*, 3025-3034. doi: 10.1093/jac/dkx264

49. Smet A.; Rasschaert G.; Martel A.; Persoons D.; Dewulf J.; Butaye P.; Catry B.; Haesebrouck F.; Herman L.; Heyndrickx M. *In situ* ESBL conjugation from avian to human *Escherichia coli* during cefotaxime administration. *J Appl Microbiol.* **2011** *110(2)*, 541-9. doi: 10.1111/j.1365-2672.2010.04907.x

50. Ludden C.; Decano A.G.; Jamrozy D.; Pickard D.; Morris D.; Parkhill J.; Peacock S.J.; Cormican M.; Downing T. Genomic surveillance of *Escherichia coli* ST131 identifies local expansion and serial replacement of subclones. *Microb Genom.* **2020** doi: 10.1099/mgen.0.000352

51. McNally, A.; et al. Diversification of Colonization Factors in a Multidrug-Resistant *Escherichia coli* Lineage Evolving under Negative Frequency-Dependent Selection. *MBio*. **2019** *10(2)*, e00644–19, https://doi.org/10.1128/mBio.00644-19.

52. Decano, A.G.; Ludden C.; Feltwell T.; Judge K.; Parkhill J.; Downing T. Complete assembly of *Escherichia coli* ST131 genomes using long reads demonstrates antibiotic resistance gene variation within diverse plasmid and chromosomal contexts. *mSphere* **2019** *4(3)*, e00130–19, https://doi.org/10.1128/mSphere.00130-19.

53. Ny S.; Sandegren L.; Salemi M.; Giske C.G. Genome and plasmid diversity of Extended-Spectrum beta-Lactamase-producing *Escherichia coli* ST131 – tracking phylogenetic trajectories with Bayesian inference. *Scientific Reports* **2019** *9*, 10291, https://doi.org/10.1038/s41598-019-46580-3.

54. Kallonen, T.; et al. Systematic longitudinal survey of invasive *Escherichia coli* in England demonstrates a stable population structure only transiently disturbed by the emergence of ST131. *Genome Res*. **2017** 27, 1437–1449, https://doi.org/10.1101/gr.216606.116.

55. Johnson J.R.; et al. Household Clustering of *Escherichia coli* Sequence Type 131 Clinical and Fecal Isolates According to Whole Genome Sequence Analysis. *Open Forum Infect Dis.* **2016** *3(3)*, ofw129.

56. Carattoli A. Resistance plasmid families in *Enterobacteriaceae*. *Antimicrob Agents Chemother.* **2009** *53(6)*, 2227-38. doi: 10.1128/AAC.01707-08

57. Nicolas-Chanoine M.H.; et al. Intercontinental emergence of *Escherichia coli* clone O25:H4-ST131 producing CTX-M-15. *J Antimicrob Chemother.* **2008** *61(2)*, 273-81.

58. Nicolas-Chanoine M.H.; Bertrand X.; Madec J;Y. *Escherichia coli* ST131, an intriguing clonal group. Clin Microbiol Rev. **2014** *27(3)*, 543-74. doi: 10.1128/CMR.00125-13

59. Woodford N.; Carattoli A.; Karisik E.; Underwood A.; Ellington M.J.; Livermore D.M. Complete nucleotide sequences of plasmids pEK204, pEK499, and pEK516, encoding CTX-M enzymes in three major *Escherichia coli* lineages from the United Kingdom, all belonging to the international O25:H4-ST131 clone. *Antimicrob. Agents Chemother.* **2009** *53(10)*, 4472-4482.

60. Phan M.D.; et al. Molecular characterization of a multidrug resistance IncF plasmid from the globally disseminated *Escherichia coli* ST131 clone. *PLoS One*. **2015** *10(4)*, e0122369. doi: 10.1371/journal.pone.0122369

61. Decano A.G.; Downing T. An *Escherichia coli* ST131 pangenome atlas reveals population structure and evolution across 4,071 isolates. *Sci Rep*. **2019** *9(1)*, 17394. doi: 10.1038/s41598-019-54004-5

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 May 2020 doi:10.20944/preprints202005.0140.v1

62. Goswami C.; Fox S.; Holden M.T.G.; Connor M.; Leanord A.; Evans T.J. Origin, maintenance and spread of antibiotic resistance genes within plasmids and chromosomes of bloodstream isolates of *Escherichia coli*. *Microb Genom*. **2020** doi: 10.1099/mgen.0.000353

63. Jørgensen S.B.; Søraas A.; Sundsfjord A.; Liestøl K.; Leegaard T.M.; Jenum P.A. Fecal carriage of extended spectrum β-lactamase producing *Escherichia coli* and *Klebsiella pneumoniae* after urinary tract infection - A three year prospective cohort study. PLoS One. 2017 12(3):e0173510. doi: 10.1371/journal.pone.0173510

64. Coyne M.J.; Zitomersky N.L.; McGuire A.M.; Earl A.M.; Comstock L.E. Evidence of extensive DNA transfer between bacteroidales species within the human gut. *MBio* **2014** *5*, e01305–14.

65. Zhao, S.; et al. Adaptive evolution within the gut microbiome of individual people. *BioRxiv* **2017** doi: 10.1101/208009

66. Bishara, A.; et al. Strain-resolved microbiome sequencing reveals mobile elements that drive bacterial competition on a clinical timescale. BioRxiv **2017** doi:10.1101/125211

67. Knudsen P.K.; et al. Transfer of a *bla*CTX-M-1-carrying plasmid between different *Escherichia coli* strains within the human gut explored by whole genome sequencing analyses. *Sci Rep*. **2018** *8(1)*, 280. doi: 10.1038/s41598-017-18659-2

68. Göttig S.; Gruber T.M.; Stecher B; Wichelhaus T.A.; Kempf V.A. In vivo horizontal gene transfer of the carbapenemase OXA-48 during a nosocomial outbreak. *Clin Infect Dis.* **2015** *60(12)*, 1808-15. doi: 10.1093/cid/civ191

69. Willemsen I.; van Esser J.; Kluytmans-van den Bergh M.; Zhou K.; Rossen J.W.; Verhulst C.; Verduin K.; Kluytmans J. Retrospective identification of a previously undetected clinical case of OXA-48-producing *K. pneumoniae* and *E. coli*: the importance of adequate detection guidelines. *Infection*. **2016** *44(1)*, 107-10. doi: 10.1007/s15010-015-0805-7

70. Evans D.R.; et al. Systematic detection of horizontal gene transfer across genera among multidrug-resistant bacteria in a single hospital. *Elife* **2020** doi: 10.7554/eLife.53886

71. Trobos M.; Lester C.H.; Olsen J.E.; Frimodt-Møller N.; Hammerum A.M. Natural transfer of sulphonamide and ampicillin resistance between *Escherichia coli* residing in the human intestine. *J. Antimicrob. Chemother.* **2009** *63*, 80–86. 10.1093/jac/dkn437

72. Karami N.; Martner A.; Enne V.I.; Swerkersson S.; Adlerberth I.; Wold A.E. Transfer of an ampicillin resistance gene between two *Escherichia coli* strains in the bowel microbiota of an infant treated with antibiotics. *J Antimicrob Chemother*. **2007** *60(5)*, 1142-5.

73. Warnes S.L.; Highmore C.J.; Keevil C.W. Horizontal transfer of antibiotic resistance genes on abiotic touch surfaces: implications for public health. *mBio*. **2012** *3(6)*, e00489-12. doi: 10.1128/mBio.00489-12.

74. Gibson M.K.; et al. Development dynamics of the preterm infant gut microbiota and antibiotic resistome. *Nature Microbiology* **2016** *1(4)*, 16024. doi: 10.1038/nmicrobiol.2016.24

75. Andersson P.; Engberg I.; Lidin-Janson G.; Lincoln K.; Hull R.; Hull S.; Svanborg C. Persistence of *Escherichia coli* bacteriuria is not determined by bacterial adherence. *Infect Immun.* **1991** *59(9)*, 2915-21.

76. Hunt M.; Mather A.E.; Sánchez-Busó L.; Page A.J.; Parkhill J.; Keane J.A.; Harris S.R. ARIBA: rapid antimicrobial resistance genotyping directly from sequencing reads. *Microb Genom.* **2017** *3(10)*, e000131. doi: 10.1099/mgen.0.000131

77. Jia; et al. CARD 2017: expansion and model-centric curation of the comprehensive antibiotic resistance database. *Nucleic Acids Research.* **2017** *45(D1)*, D566-D573. doi: 10.1093/nar/gkw1004

78. Rowe W.P.M.; Winn M.D. Indexed variation graphs for efficient and accurate resistome profiling. *Bioinformatics*. **2018** *34(21)*, 3601-3608. doi: 10.1093/bioinformatics/bty387

79. Phan M.D.; Bottomley A.L.; Peters K.M.; Harry E.J.; Schembri M.A. Uncovering novel susceptibility targets to enhance the efficacy of third-generation cephalosporins against ESBL-producing uropathogenic *Escherichia coli*. *J Antimicrob Chemother*. **2020** *dkaa023* doi: 10.1093/jac/dkaa023

80. Merlin T.L.; Davis G.E.; Anderson W.L.; Moyzis R.K.; Griffith J.K. Aminoglycoside uptake increased by tet gene expression. *Antimicrob Agents Chemother*. **1989** 33(9):1549-52

81. Stavropoulos T.A.; Strathdee C.A. Expression of the tetA(C) tetracycline efflux pump in *Escherichia coli* confers osmotic sensitivity. FEMS Microbiol. Lett. **1000** 190, 147–150.

82. Heng J, Zhao Y, Liu M, Liu Y, Fan J, Wang X, Zhao Y, Zhang XC. Substrate-bound structure of the *E. coli* multidrug resistance transporter MdfA. *Cell Res*. **2015** *25(9)*, 1060-73. doi: 10.1038/cr.2015.94

83. Nguyen-Distèche M.; Fraipont C.; Buddelmeijer N.; Nanninga N. The structure and function of *Escherichia coli* penicillin-binding protein 3. *Cell Mol Life Sci*. **1998** *54(4)*, 309-16.

84. Nakamura M.; Maruyama I.N.; Soma M.; Kato J.; Suzuki H.; Horota Y. On the process of cellular division in *Escherichia coli*: nucleotide sequence of the gene for penicillin-binding protein 3. *Mol Gen Genet*. **1983** 191(1):1-9

85. Sun J.; Deng Z.; Yan A. Bacterial multidrug efflux pumps: mechanisms, physiology and pharmacological exploitations. *Biochem Biophys Res Commun*. **2014** *453(2)*, 254-67. doi: 10.1016/j.bbrc.2014.05.090

86. Curtis N.A.; Eisenstadt R.L.; Turner K.A.; White A.J. Inhibition of penicillin-binding protein 3 of *Escherichia coli* K-12. Effects upon growth, viability and outer membrane barrier function. *J Antimicrob Chemother*. **1985** *16(3)*, 287-96

87. Toh H., et al. Complete genome sequence of the wild-type commensal *Escherichia coli* strain SE15, belonging to phylogenetic group B2. *Journal of Bacteriology* **2010** *192(4)*, 1165-1166.

88. Johnson J.R.; Porter S.; Thuras P.; Castanheira M. Epidemic Emergence in the United States of *Escherichia coli* Sequence Type 131-H30 (ST131-H30), 2000 to 2009. *Antimicrob Agents Chemother*. **2017** *61(8)*, e00732-17. doi: 10.1128/AAC.00732-17

89. Karisik E.; Ellington M.J.; Pike R.; Warren R.E.; Livermore D.M.; Woodford N. Molecular characterization of plasmids encoding CTX-M-15 beta-lactamases from *Escherichia coli* strains in the United Kingdom. *J Antimicrob Chemother*. **2006** *58(3)*, 665-8.

90. Dudley E.G.; Abe C.; Ghigo J.M.; Latour-Lambert P.; Hormazabal J.C.; Nataro J.P. An IncI1 plasmid contributes to the adherence of the atypical enteroaggregative *Escherichia coli* strain C1096 to cultured cells and abiotic surfaces. *Infect Immun*. **2006** *74(4)*, 2102-14

91. Dhanji H.; Doumith M.; Hope R.; Livermore D.M.; Woodford N. IS*Ecp1*-mediated transposition of linked *bla*CTX-M-3 and *bla*TEM-1b from the IncI1 plasmid pEK204 found in clinical isolates of *Escherichia coli* from Belfast, UK. *J Antimicrob Chemother*. **2011** *66(10)*, 2263-5. doi: 10.1093/jac/dkr310

92. Runyen-Janecky L.J.; Hong M.; Payne S.M. The virulence plasmid-encoded impCAB operon enhances survival and induced mutagenesis in *Shigella flexneri* after exposure to UV radiation. *Infect Immun. 67(3)*, 1415-23.

93. Akiba M.; et al. Distribution and Relationships of Antimicrobial Resistance Determinants among Extended-Spectrum-Cephalosporin-Resistant or Carbapenem-Resistant *Escherichia coli* Isolates from Rivers and Sewage Treatment Plants in India. *Antimicrob. Agents Chemother*. **2016** *60(5)*, 2872-2980.

94. Carattoli A.; et al. *In Silico* Detection and Typing of Plasmids using PlasmidFinder and Plasmid Multilocus Sequence Typing. 2014. *Antimicrob. Agents Chemother.* **2014** *58(7)*, 3895-3903. doi: 10.1128/AAC.02412-14.

95. Purves K.; Macintyre L.; Brennan D.; Hreggviðsson G.; Kuttner E.; Ásgeirsdóttir M.; Young L.; Green D.; Edrada-Ebel R.; Duncan K. Using molecular networking for microbial secondary metabolite bioprospecting. *Metabolites* **2016** 6(1), 2.

96. Typas A.; Sourjik V. Bacterial protein networks: properties and functions. *Nat Rev Microbiol*. **2015** *13(9)*, 559-72. doi: 10.1038/nrmicro3508

97. Kirschner M.; Gerhart J. Evolvability. Proc Natl Acad Sci USA. **1998** *95(15)*, 8420-8427; https://doi.org/10.1073/pnas.95.15.8420

98. Szklarczyk D.; et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* **2019** *47(D1)*, D607-D613. doi: 10.1093/nar/gky1131

99. Miryala S.K.; Ramaiah S. Exploring the multi-drug resistance in *Escherichia coli* O157:H7 by gene interaction network: A systems biology approach. *Genomics*. **2019** *111(4)*, 958-965. doi: 10.1016/j.ygeno.2018.06.002

100. Zong Z. Complete sequence of pJIE186-2, a plasmid carrying multiple virulence factors from a sequence type 131 *Escherichia coli* O25 strain. *Antimicrob Agents Chemother.* **2013** *57(1)*, 597-600. doi: 10.1128/AAC.01081-12

101. Clark G.; Paszkiewicz K.; Hale J.; Weston V.; Constantinidou C.; Penn C.; Achtman M.; McNally A. Genomic analysis uncovers a phenotypically diverse but genetically homogeneous *Escherichia coli* ST131 clone circulating in unrelated urinary tract infections. *J Antimicrob Chemother.* **2012** *67(4)*, 868-77. doi: 10.1093/jac/dkr585

102. Holden N.; Totsika M.; Dixon L.; Catherwood K.; Gally D.L. Regulation of P-fimbrial phase variation frequencies in *Escherichia coli* CFT073. Infect Immun. **2007** *75(7)*, 3325-34.

103. Spaulding C.N.; et al. Selective depletion of uropathogenic *E. coli* from the gut by a FimH antagonist. Nature. **2017** *546(7659)*, 528-532. doi: 10.1038/nature22972

104. Xia Y.; Gally D.; Forsman-Semb K.; Uhlin B.E. Regulatory cross-talk between adhesin operons in *Escherichia coli*: inhibition of type 1 fimbriae expression by the PapB protein. *EMBO J* **2000** *19*, 1450 –1457. https://doi.org/10.1093/emboj/19.7.1450.

105. Snyder J.A.; Haugen B.J.; Lockatell C.V.; Maroncle N.; Hagan E.C.; Johnson D.E.; Welch R.A.; Mobley H.L. 2005. Coordinate expression of fimbriae in uropathogenic *Escherichia coli*. *Infect Immun 73*, 7588 –7596. https://doi.org/10.1128/IAI.73.11.7588-7596.2005.

106. Forsman K.; Göransson M.; Uhlin B.E. 1989. Autoregulation and multiple DNA interactions by a transcriptional regulatory protein in *E. coli* pili biogenesis. *EMBO J* **1989** *8*, 1271–1277.

107. Holden N.J.; et al. Demonstration of regulatory cross-talk between P fimbriae and type 1 fimbriae in uropathogenic *Escherichia coli*. *Microbiology*. **2006** *152(Pt4)*, 1143-53.

108. Brzuszkiewicz E, et al. How to become a uropathogen: comparative genomic analysis of Extraintestinal pathogenic *Escherichia coli* strains. Proc Natl Acad Sci U S A. **2006** *103(34)*, 12879-84.

109. Welch R.A.; et al. Extensive mosaic structure revealed by the complete genome sequence of uropathogenic *Escherichia coli*. *Proc Natl Acad Sci U S A.* **2002** 99(26):17020-4

110. Lillington J.; Geibel S.; Waksman G. Biogenesis and adhesion of type 1 and P pili. *Biochim Biophys Acta.* **2014** *1840(9)*, 2783-93. doi: 10.1016/j.bbagen.2014.04.021

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 May 2020 doi:10.20944/preprints202005.0140.v1

111. Forsberg K.J.; Reyes A.; Wang B.; Selleck E.M.; Sommer M.O.; Dantas G. The shared antibiotic resistome of soil bacteria and human pathogens. *Science*. **2012** *337(6098)*, 1107-11. doi: 10.1126/science.1220761

112. Chen S.L.; et al. Identification of genes subject to positive selection in uropathogenic strains of *Escherichia coli*: a comparative genomics approach. *Proc Natl Acad Sci U S A*. **2006** *103(15)*, 5977-82.

113. Tenaillon O.; Skurnik D.; Picard B.; Denamur E. The population genetics of commensal *Escherichia coli*. *Nat Rev Microbiol*. **2010** *8(3)*, 207-17. doi: 10.1038/nrmicro2298

114. Didelot X.; Méric G.; Falush D.; Darling A.E. Impact of homologous and non-homologous recombination in the genomic evolution of *Escherichia coli*. *BMC Genomics*. **2012** *13*, 256. doi: 10.1186/1471-2164-13-256.

115. Dixit P.D.; Pang T.Y.; Studier F.W.; Maslov S. Recombinant transfer in the basic genome of *Escherichia coli*. *Proc Natl Acad Sci U S A*. **2015** *112(29)*, 9070-5. doi: 10.1073/pnas.1510839112

116. Tchesnokova V.; Radey M.; Chattopadhyay S.; Larson L.; Weaver J.L.; Kisiela D.; Sokurenko E.V. Pandemic fluoroquinolone resistant *Escherichia coli* clone ST1193 emerged via simultaneous homologous recombinations in 11 gene loci. *Proc Natl Acad Sci U S A*. **2019** *116(29)*, 14740-14748. doi: 10.1073/pnas.1903002116

117. Smillie C.S.; Smith M.B.; Friedman J.; Cordero O.X.; David L.A.; Alm E.J. Ecology drives a global network of gene exchange connecting the human microbiome. *Nature*. **2011** *480(7376)*, 241-4. doi: 10.1038/nature10571.

118. Lloyd-Price J.; et al. Strains, functions and dynamics in the expanded Human Microbiome Project. Nature **2017** *550(7674)*, 61-66.

119. Tamburini F.B.; Andermann T.M.; Tkachenko E.; Senchyna F.; Banaei N.; Bhatt A.S. Precision identification of diverse bloodstream pathogens in the gut microbiome. *Nat Med*. **2018** *24(12)*, 1809-1814. doi: 10.1038/s41591-018-0202-8

120. Sundén F.; Håkansson L.; Ljunggren E.; Wullt B. *Escherichia coli* 83972 bacteriuria protects against recurrent lower urinary tract infections in patients with incomplete bladder emptying. *J Urol*. **2010** *184(1)*, 179-85. doi: 10.1016/j.juro.2010.03.024

121. Sundén F.; Håkansson L.; Ljunggren E.; Wullt B. Bacterial interference—is deliberate colonization with *Escherichia coli* 83972 an alternative treatment for patients with recurrent urinary tract infection? *Int J Antimicrob Agents*. **2006** *28 Suppl 1*, S26-9.

122. Watts R.E.; et al. *Escherichia coli* 83972 expressing a P fimbriae oligosaccharide receptor mimic impairs adhesion of uropathogenic *E. coli*. *J Infect Dis*. **2012** 206(8), 1242-9.

123. Zdziarski J.; et al. Host imprints on bacterial genomes--rapid, divergent evolution in individual patients. *PLoS Pathog*. **2010** *6(8)*, e1001078. doi: 10.1371/journal.ppat.1001078

124. Duprilot M.; et al. Success of *Escherichia coli* O25b:H4 ST131 clade C associated with a decrease in virulence. BioRxiv **2019** doi: https://doi.org/10.1101/786350.

125. Mora A.; et al. Virulence patterns in a murine sepsis model of ST131 *Escherichia coli* clinical isolates belonging to serotypes O25b:H4 and O16:H5 are associated to specific virotypes. *PLoS ONE* **2014** *9*, e87025.

126. Fondi M.; Karkman A.; Tamminen M.V.; Bosi E.; Virta M.; Fani R.; Alm E.; McInerney J.O. Every Gene Is Everywhere but the Environment Selects: Global Geolocalization of Gene Sharing in Environmental Samples through Network Analysis. *Genome Biol Evol*. **2016** *8(5)*, 1388-400. doi: 10.1093/gbe/evw077

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 May 2020 doi:10.20944/preprints202005.0140.v1

127. Downing T. Tackling Drug Resistant Infection Outbreaks of Global Pandemic *Escherichia coli* ST131 Using Evolutionary and Epidemiological Genomics. *Microorganisms*. **2015** *3(2)*, 236-67. doi: 10.3390/microorganisms3020236

128. Leggett R.M.; et al. Rapid MinION profiling of preterm microbiota and antimicrobial-resistant pathogens. *Nat Microbiol* **2020** *5*, 430–442 https://doi.org/10.1038/s41564-019-0626-z.

129. Smith H.; et al. Characterization of epidemic IncI1-Iγ plasmids harboring Ambler class A and C genes in *Escherichia coli* and *Salmonella enterica* from animals and humans. *Antimicrob Agents Chemother* **2015** *59*, 5357-5365. https://doi.org/10.1128/AAC.05006-14.

130. Rozwandowicz M, Brouwer MSM, Fischer J, Wagenaar JA, Gonzalez-Zorn B, Guerra B, Mevius DJ, Hordijk J. 2018. Plasmids carrying antimicrobial resistance genes in Enterobacteriaceae. J Antimicrob Chemother 73:1121–1137. https://doi.org/10.1093/jac/dkx488

131. Rodríguez-Navarro J, Miró E, Brown-Jaque M, Hurtado JC, Moreno A, Muniesa M, González-López JJ, Vila J, Espinal P, Navarro F. Comparison of Commensal and Clinical Isolates for Diversity of Plasmids in Escherichia coli and Klebsiella pneumoniae. Antimicrob Agents Chemother. 2020 64(5). pii: e02064-19. doi: 10.1128/AAC.02064-19

132. Dimitriu T.; Medaney F.; Amanatidou E.; Forsyth J.; Ellis R.J.; Raymond B. Negative frequency dependent selection on plasmid carriage and low fitness costs maintain extended spectrum β-lactamases in *Escherichia coli. Sci Rep*. **2019** *9(1)*, 17211. doi: 10.1038/s41598-019-53575-7

133. Totsika M.; et al. Insights into a multidrug resistant *Escherichia coli* pathogen of the globally disseminated ST131 lineage: genome analysis and virulence mechanisms. *PLoS One* **2011** *6*, e26578.

134. Schembri M.A.; Zakour N.L.; Phan M.D.; Forde B.M.; Stanton-Cook M.; Beatson S.A. Molecular Characterization of the Multidrug Resistant *Escherichia coli* ST131 Clone. *Pathogens*. **2015** *4(3)*, 422-30. doi: 10.3390/pathogens4030422

135. Wurpel, D.J.; et al. Comparative proteomics of uropathogenic *Escherichia coli* during growth in human urine identify UCA-like (UCL) fimbriae as an adherence factor involved in biofilm formation and binding to uroepithelial cells. *J. Proteomics* **2016** *131*, 177–189.

136. Lane M.C.; Mobley H.L. Role of P-fimbrial-mediated adherence in pyelonephritis and persistence of uropathogenic *Escherichia coli* (UPEC) in the mammalian kidney. *Kidney Int* **2007** *72*, 19 –25. https://doi.org/10.1038/sj.ki.5002230.

137. Gunther, N.W.; Lockatell, V.; Johnson, D.E.; Mobley, H.L.T. *In vivo* dynamics of type, 1 fimbria regulation in uropathogenic *Escherichia coli* during experimental urinary tract infection. Infect Immun **2001** 69, 2838-2846.

138. Holden N.J.; Uhlin B.E.; Gally D.L. PapB paralogues and their effect on the phase variation of type 1 fimbriae in *Escherichia coli*. *Mol Microbiol.* **2001** *42(2)*, 319-30

139. Holden N.J.; Gally D. Switches, cross-talk and memory in *Escherichia coli* adherence. *J Med Microbiol*. **2004** *53(Pt 7)*, 585-593. doi: 10.1099/jmm.0.05491-0

140. Whitmer G.R.; Moorthy G.; Arshad M. The pandemic *Escherichia coli* sequence type 131 strain is acquired even in the absence of antibiotic exposure. *PLoS Pathog*. **2019** *15(12)*, e1008162. doi: 10.1371/journal.ppat.1008162

141. NIH HMP Working Group; et al. The NIH Human Microbiome Project. *Genome Res.* **2009** *19(12)*, 2317-23. doi: 10.1101/gr.096651.109.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 May 2020 doi:10.20944/preprints202005.0140.v1

142. Rudick C.N.; Taylor A.K.; Yaggie R.E.; Schaeffer A.J.; Klumpp D.J. Asymptomatic Bacteriuria *Escherichia coli* are Live Biotherapeutics for UTI. *PLoS One* **2014** *9(11)*, 1-9.

143. Hancock V.; Seshasayee A.S.; Ussery D.W.; Luscombe N.M.; Klemm P. Transcriptomics and adaptive genomics of the asymptomatic bacteriuria Escherichia *coli* strain 83972. *Mol Genet Genomics.* **2008** *279(5)*, 523-34. doi: 10.1007/s00438-008-0330-9.

144. Floyd, R.V.; Upton, M.; Hultgren, S.J.; Wray, S.; Burdyga, T.V.; Winstanley, C. *Escherichia coli*-mediated impairment of ureteric contractility is uropathogenic *E. coli* specific. *J. Infect. Dis.* **2012**, *206*, 1589–1596.

145. Bolger A.M.; Lohse M.; Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. **2014** *30(15)*, 2114-20. doi: 10.1093/bioinformatics/btu170

146. Li H.; et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. **2009** *25(16)*, 2078-9. doi: 10.1093/bioinformatics/btp352

147. Gupta S.K.; et al. ARG-ANNOT, a new bioinformatic tool to discover antibiotic resistance genes in bacterial genomes. *Antimicrob Agents Chemother.* **2014** *58(1)*, 212-20. doi: 10.1128/AAC.01310-13

148. Zankari E.; et al. Identification of acquired antimicrobial resistance genes. *J Antimicrob Chemother.* **2012** *67(11)*, 2640-4. doi: 10.1093/jac/dks261

149. Nicolas-Chanoine M.H.; et al. The ST131 *Escherichia coli* H22 subclone from human intestinal microbiota: Comparison of genomic and phenotypic traits with those of the globally successful H30 subclone. *BMC Microbiol.* **2017** *17(1)*, 71. doi: 10.1186/s12866-017-0984-8

150. Jones P.; et al. InterProScan 5: genome-scale protein function classification. Bioinformatics. **2014** *30(9)*, 1236-40. doi: 10.1093/bioinformatics/btu031

151. Carver T.; Berriman M.; Tivey A.; Patel C.; Böhme U.; Barrell B.G.; Parkhill J.; Rajandream M.A. Artemis and ACT: viewing, annotating and comparing sequences stored in a relational database. Bioinformatics. **2008** *24(23)*, 2672-6. doi: 10.1093/bioinformatics/btn529.

152. Sievers F.; et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology* **2011** *7*, 539 doi:10.1038/msb.2011.75

153. Nicolau M.; Levine A.J.; Carlsson G. Topology based data analysis identifies a subgroup of breast cancers with a unique mutational profile and excellent survival. *Proc Natl Acad Sci U S A.* **2011** *108(17)*, 7265-70. doi: 10.1073/pnas.1102826108

154. Vietoris L. Über den höheren Zusammenhang kompakter Räume und eine Klasse von zusammenhangstreuen Abbildungen. *Mathematische Annalen* **1927** *97(1)*, 454–472.

155. Rahm A. HomologyLive. **2019** http://math.uni.lu/~rahm/HomologyLive.html.

156. LinBox, The LinBox Group. Exact Linear Algebra over the Integers and Finite Rings, Version 1.1.6, **2008**.

157. Kim S.R.; Komano T. The plasmid R64 thin pilus identified as a type IV pilus. *J Bacteriol*. **1997** *179(11)*, 3594-603.