

## **Title**

Functionally redundant forms of extended-spectrum beta-lactamases and aminoglycoside-modifying enzymes drive the evolution of two distinct multidrug resistance gene clusters in clinical populations of EXPEC

Jay W. Kim<sup>1,2</sup>, Portia Mira<sup>3</sup>, Patricia P. Chan<sup>2,4</sup>, Todd M. Lowe<sup>2</sup>, Miriam Barlow<sup>3\*</sup>, Manel Camps<sup>1\*</sup>

\*Co-corresponding authors

<sup>1</sup>Department of Microbiology and Environmental Toxicology, University of California at Santa Cruz, Santa Cruz, CA, 95064, USA

<sup>2</sup>Department of Biomolecular Engineering, University of California at Santa Cruz, Santa Cruz, CA, 95064, USA

<sup>3</sup>School of Natural Sciences, University of California at Merced, 5200 North Lake Road, Merced, CA, 95343, USA

<sup>4</sup>Maverix Biomics Inc., Los Altos, CA 94022, USA

## **Correspondent emails**

MB: [miriam.barlow@gmail.com](mailto:miriam.barlow@gmail.com)

MC: [mcamps@ucsc.edu](mailto:mcamps@ucsc.edu)

Running title: Redundant resistance gene clusters in EXPEC

Keywords: antibiotic resistance, epidemiology, plasmid, horizontal gene transfer, conjugation, antibiotic resistance spread, high risk clones, MLST typing, IncF, epidemic clones

Abbreviations: ESBL=Extended Spectrum Beta-Lactamase, CG=complementarity group, HGT=horizontal gene transfer, Inc=plasmid incompatibility, AAC=aminoglycoside acetyl transferase

## Abstract

We evaluate the distribution of genetic markers for antibiotic resistance in 276 genomic sequences of Extraintestinal Pathogenic *E. coli* from two hospitals on the U.S. West coast. Plasmid-borne genes encoding drug-inactivating enzymes dominate the distribution of aminoglycoside and  $\beta$ -lactam resistance markers. These genes can be assigned based on their distribution to two mutually exclusive complementarity groups (CGs: CG1 and CG2) with each displaying genetic linkage and minimal functional overlap. CG1 includes genes encoding OXA-1 and AAC(6')-Ib-cr, frequently also CTX-M-15, and sometimes AAC(3)-IIe.2, a variant of AAC(3)-IIe; CG2 includes AAC(3)-IIId tightly linked to TEM-1, and occasionally also to genes encoding CTX-M-14-like  $\beta$ -lactamases. This binary distribution of aminoglycoside and  $\beta$ -lactamase resistance genes suggests a convergence between two different evolutionary solutions, and results in a ubiquitous functional redundancy in the clinical populations. CG1 and CG2 are largely carried in IncF plasmids, of which we distinguish seven classes based on Rpt-A1 sequence homology. Both CG1 and CG2 genes are found in two different IncF plasmid classes, demonstrating their pervasive mobility across plasmid backbones. Different CG genes and IncF plasmid classes are found in a wide range of MLSTs, highlighting the prevalence of horizontal gene transfer. We also identify at least five clonally expanding MLSTs, which represent high-risk clones: ST131, ST95, ST73, ST127, and ST69. The identification of clonally-expanding types, the discovery of CGs that are ubiquitously spread in diverse clinical strains, and the functional redundancy that these two groups represent have significant implications for monitoring and controlling the spread of resistance.

## Introduction

Antibiotic resistance represents a growing threat to public health, increasing medical costs worldwide. Particularly, the spread of multidrug-resistant *E. coli* clones is a serious public health

concern because of its symbiotic relationship with humans and animals (Li et al. 2007; Peirano et al. 2012), and because of its frequent implication in opportunistic infection of its hosts, causing urinary tract infections, sepsis and wound infections both in clinical and community-associated settings (Robicsek et al. 2006; Price et al. 2013; Yano et al. 2013). The spread of resistance is further accelerated by incorrect or unnecessary antibiotic treatments (Ventola 2015; Burnham et al. 2017). Thus, understanding the mechanisms driving the spread of resistance and improving the accuracy of therapeutic decisions through point-of-care diagnostics should be of great help to both slow down the development of antibiotic resistance and to reduce the costs associated with antibiotic-resistant infections.

Here, we evaluate the distribution of known antibiotic resistance genes in 276 genomic sequences of Extraintestinal Pathogenic *E. coli* (ExPEC) isolated in two West Coast, U.S. hospitals: University of Washington (UW) in Seattle, Washington (Salipante et al. 2015), and the Dignity Health Mercy Medical Center (DHMMC), in Merced, California. We focus on resistance to the three classes of antibiotics most frequently used for the treatment of ExPEC infections and of enterobacterial infections in general, namely aminoglycosides,  $\beta$ -lactams, and fluoroquinolones (Rodriguez-Bano et al. 2018) (for a list of antibiotics included in this study and their distribution in the population see **Methods**).

Aminoglycosides are a large class of structurally related compounds that inhibit different steps of bacterial protein synthesis (Magnet and Blanchard 2005; Jana and Deb 2006; Shakil et al. 2008). Resistance can result from direct modification and inactivation of the drug molecule by aminoglycoside-modifying enzymes (Ramirez and Tolmasky 2010), methylation of the amino acyl site of 16S rRNA through 16S rRNA methyltransferases (Doi et al. 2016), and by mutations in genes encoding ribosomal proteins (e.g. *rpsL*). Additional contributors to aminoglycoside



resistance include altered expression and/or mutations in outer membrane porins, two-component systems and efflux systems (Cag et al. 2016).

$\beta$ -lactam antibiotics include penicillins, cephalosporins, monobactams and carbapenems. These drugs diffuse into the cells' periplasm, where they bind to and inactivate penicillin-binding proteins, which are essential for cell wall maintenance, resulting in cell death (Zapun et al. 2008; Sauvage and Terrak 2016). Resistance to  $\beta$ -lactams is largely due to hydrolysis by  $\beta$ -lactamases, which in ExPEC include TEM, SHV, CTX-M, OXA, AmpC, and a variety of carbapenemases (Gutkind et al. 2013). All of these proteins have activity against penicillins. Some of them can also protect against oxyimino (3<sup>rd</sup> and 4<sup>th</sup> generation) cephalosporins and against monobactams, a phenotype known as Extended-Spectrum  $\beta$ -Lactamase (ESBL) resistance (Ghafourian et al. 2015). ESBL activity can be constitutive (CTX-Ms) or evolved, as is the case of mutant TEM, mutant SHVs, and some OXAs. Carbapenems have been introduced more recently for the treatment of ESBL resistance, and only carbapenemases (KPC, NDM, VIM) (Martinez-Martinez and Gonzalez-Lopez 2014; Jeon et al. 2015) and most  $\beta$ -lactamases of the OXA family exhibit activity against them (Mairi et al. 2017).

Fluoroquinolones inhibit type-II topoisomerase activity, blocking replication. Resistance frequently derives from mutations in the two main topoisomerase II-type enzymes present in *E. coli*, namely DNA gyrase and DNA topoisomerase IV. Each consists of two subunits: GyrA and GyrB for DNA gyrase and ParC and ParE for DNA topoisomerase IV. In *E. coli* and other Gram negative pathogens, DNA gyrase is the preferential target for fluoroquinolones but highly-resistant organisms tend to contain resistance-conferring mutations in both enzymes (Redgrave et al. 2014). These mutations tend to occur specifically in the quinolone resistance determining

regions (QRDR) of GyrA and ParC, preventing the binding of the drug to these targets (Redgrave et al. 2014).

Thus, a considerable number of drug resistance factors have already been identified. The emerging picture is that antibiotic resistance involves sequential lines of defense targeting drug influx, accumulation, target binding, or downstream toxicity (Yelin and Kishony 2018). We are also beginning to understand the processes leading to the acquisition and spread of resistance genes in the population, but significant gaps remain due both to the complex distribution of pathogens within hospitals and across the community (Peirano et al. 2012; Kang et al. 2013), and also due to the difficulty in adequately detecting resistance genes by PCR or molecular hybridization methods because these genes can exhibit a high degree of genetic variability (Leinberger et al. 2010; Rubtsova et al. 2010). With the advent of next-generation sequencing, the application of WGS technologies in clinical diagnostics and research has been a fast-growing trend, particularly in tracking hospital outbreaks of bacterial pathogens (Snitkin et al. 2012; Luo et al. 2014; Fitzpatrick et al. 2016), and predicting drug resistance based on bacterial sequence rather than with phenotypic assays (Gordon et al. 2014; Coll et al. 2015; Walker et al. 2015).

Plasmids are autonomously replicating units that greatly facilitate the spread of antibiotic resistance (Carattoli 2013; Brolund and Sandegren 2016). Most aminoglycoside-modifying enzymes and  $\beta$ -lactamases are carried in plasmids (Banerjee and Johnson 2013; Brolund and Sandegren 2016). While the main mutations conferring fluoroquinolone resistance are chromosomal, plasmid-borne proteins of the pentapeptide repeat family encoded by quinolone resistance (*qnr*) genes can also confer moderate fluoroquinolone resistance by protecting DNA gyrase and topoisomerase IV (Rodriguez-Martinez et al. 2016; Yanat et al. 2017). Finally, the

aminoglycoside acetyltransferase AAC(6')-Ib-cr can acetylate ciprofloxacin and norfloxacin, thus contributing to resistance (Briales et al. 2012).

Plasmids carry mobile elements including insertion sequences (IS), integrons, and transposition units (reviewed in (Partridge 2011; Guedon et al. 2017)) that facilitate the capture of resistance genes from a shared environmental pool (Crofts et al. 2017). Each particular resistance gene is generally immediately associated with a specific mobile element, possibly reflecting individual gene capture processes (Partridge 2011). Through recombination, these mobile elements in turn also facilitate the acquisition of additional antibiotic resistance genes through a stochastic process known as *genetic capitalism*, which favors the acquisition of additional antibiotic resistance genes by strains already exhibiting antibiotic resistance (Canton et al. 2003; Canton and Ruiz-Garbajosa 2011) and also the clustering together of these antibiotic resistance genes in large multi-resistance regions (MRRs) (Baquero 2004; Partridge et al. 2011). Plasmids also facilitate the spread of resistance through conjugation, a contact-dependent and energy-driven process mediated by a molecular machinery that is related to type IV secretion systems (Ilangovan et al. 2015; Banuelos-Vazquez et al. 2017). This means that plasmids can transfer across unrelated bacterial strains, spreading antibiotic resistance genes in the process (Shin et al. 2012; Hu et al. 2014).

The minimal portion of a plasmid replicating with the characteristic copy number of the parent is known as a replicon. Replicons are classified into incompatibility (Inc) groups based on their ability to be propagated stably in the same cell (Novick 1987). A total of twenty-seven Inc groups are known in *Enterobacteriaceae*, although their representation varies substantially in naturally occurring bacteria. IncF plasmids are one of the most widely diffused plasmid families in clinically relevant *Enterobacteriaceae*, (Johnson et al. 2007) and have played a major role in the dissemination of bla<sub>CTX-M-15</sub>, and bla<sub>CTX-M-14</sub> through *ISEcp1* capture and mobilization and through

their association with the high-risk strain ST131 (Nicolas-Chanoine et al. 2014). IncF plasmids are unusual in that they frequently contain multiple replicons, possibly resulting from the recombination of different plasmids (Villa et al. 2010). In addition, IncF plasmids have an abundance of mobile elements, which also facilitate genetic exchange. As a result, IncF plasmids show a very high level of recombination, with widespread deletions and truncations, and highly variable modules associated with virulence and biochemical pathways (Partridge et al. 2011; Li et al. 2015).

The adaptive success of *high-risk clones* (host strains) often contributes to the spread of plasmid-mediated antibiotic resistance. Sequence type 131 (ST131) is a case in point, evolving resistance to fluoroquinolones through *gyrA* and *parC* mutations (Cagnacci et al. 2008; Johnson et al. 2013), acquiring bla<sub>TEM-1</sub> and bla<sub>OXA-1</sub> IncFII plasmids, with the subsequent capture of ESBL  $\beta$ -lactamases (Woodford et al. 2011; Olesen et al. 2013), and often also capturing aminoglycoside resistance genes (Li et al. 2015).

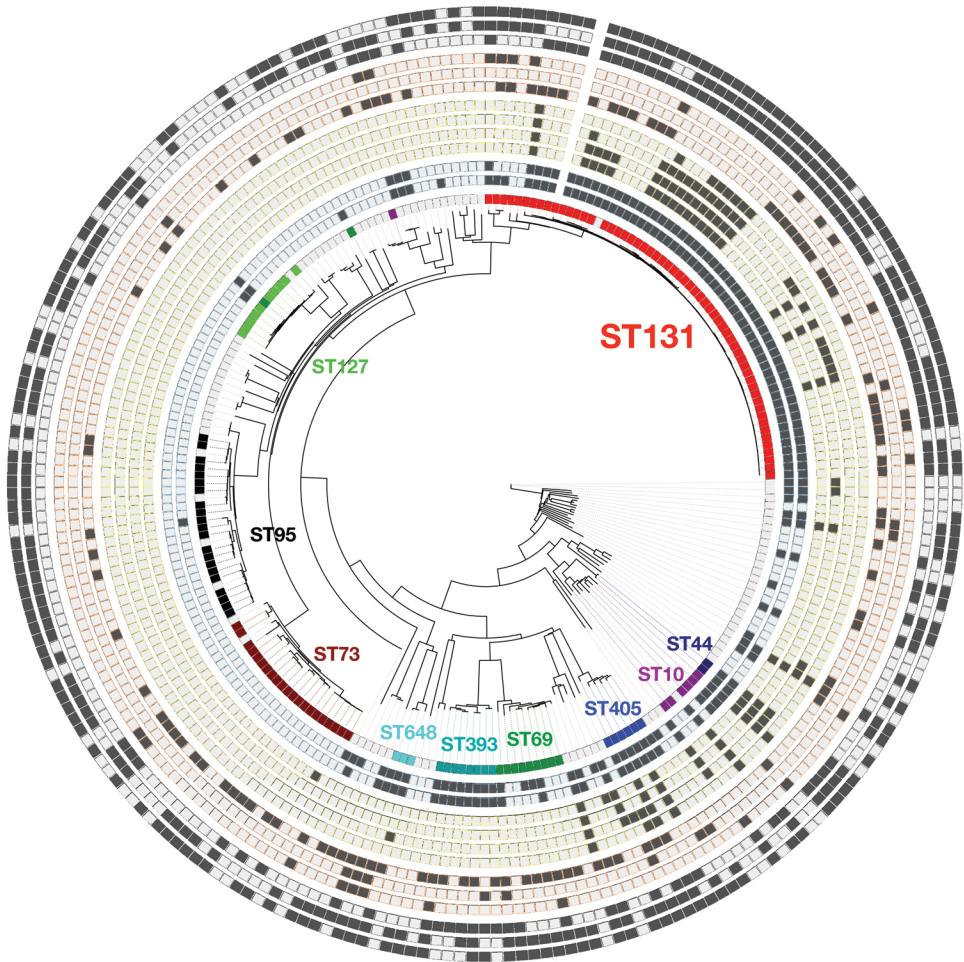
Here, we conducted a large-scale genomic study using data derived from ExPEC isolates from two U. S. West Coast hospitals to identify the distribution of known genetic antibiotic resistance genes in ExPEC. We found extensive genetic exchange associated with IncF plasmid recombination and widespread Horizontal Gene Transfer (HGT), allowing the co-existence of alternate determinants of  $\beta$ -lactam or aminoglycoside resistance in the population. Further, we find that alternate determinants of resistance are genetically linked, forming two complementarity groups. The identification of distinct complementarity groups has substantial implications for our understanding of how drug resistance evolves as well as for diagnosis and monitoring of antibiotic resistance.

## Results

We performed whole genome sequencing of 45 ESBL-resistant ExPEC isolates from Dignity Health Mercy Medical Center (DHMMC) and obtained 384 assembled draft genomes from a published University of Washington (UW) study (Salipante et al. 2015). Both of these sets had been tested for susceptibility against a panel of antibiotics, with some differences in panel composition between the two. From these, we selected 276 genomes (43 from DHMMC and 233 from UW) by removing longitudinal duplicates and controlling for sequencing quality (Details about the selection of clinical samples and the processing of the sequences are provided in *Methods* and *Supplemental Material*). The antibiotics considered in our study are shown in **Supplemental\_Table\_S1**.

### *Phylogenetic structure*

As an initial indication of strain diversity, we performed multilocus sequence typing (MLST) according to the Achtman classification (Maiden et al. 1998) (**Supplemental\_Data\_S1**). The most abundant MLSTs included ST131 (65 samples), ST95 (27 samples), ST73 (22 samples), ST127 (15 samples) and ST69 (12 samples). At the other end of the spectrum, 68 of the 276 genomes corresponded to MLSTs with 3 or less representatives. To obtain higher resolution, we generated a phylogeny based on SNP variation across the genome (**Fig. 1**).



**Figure 1. Phylogeny of 276 ExPEC strains obtained from two U.S. West Coast hospitals.**

Phylogenetic relationships among these strains are represented as a neighbor-joining tree computed based on pairwise comparisons of genomic SNPs. The Circos diagram surrounding the phylogeny shows, for each phylogenetic taxa, an MLST classification, presence of various resistance markers and plasmid replicons. Circular panels represent the following from the innermost to outermost positions. **Panel 1**, MLST classification: ST131 (red), ST44, (navy), ST10 (purple), ST405 (blue), ST69 (green), ST393 (teal), ST648 (cyan), ST73 (maroon), ST95 (black), ST127 (lime). **Panels 2 and 3** (outlined in blue), fluoroquinolone resistance mutations:

GyrA-S83X and ParC-S80I, respectively. **Panels 4-8** (outlined in yellow), complementarity group 1: genes for OXA-1, AAC(6')-Ib-cr, CTX-M-15, and AAC(3)-Ile.2, respectively. **Panels 9-11** (outlined in orange), complementarity group 2: genes for TEM-1, CTX-M-9-family and AAC(3)-IId, respectively. **Panels 12-14** (outlined in grey), IncF replicons: IncFIA, IncFIB, IncFII, respectively.

In this phylogeny, clonal expansion appears as poorly resolved clades, corresponding to the following MLSTs: ST131, ST95, ST73, ST127, and ST69. Consistent with recent expansion, these types coincide with the most well represented MLSTs. In general, the phylogeny is in good agreement with the MLST classification, with some exceptions (*Supplemental Material*).

#### *Selection of antibiotic resistance markers*

We compiled a list of genetic resistance genes reported for ExPEC and also for other Gram negatives. This list included both genes and point mutations known to be associated with the acquisition of resistance against drugs belonging to the three antibiotic classes that are most frequently used in the clinic for the treatment of enterobacterial infection, namely  $\beta$ -lactams, aminoglycosides and fluoroquinolones (Lanza et al. 2014; Zowawi et al. 2015; Cerceo et al. 2016; Rodriguez-Bano et al. 2018).

UniProt accessions and protein sequences for all known antibiotic resistance genes considered in our study are provided in **Supplemental\_Data\_S2**. Details about the drugs tested and the number of clinical isolates tested for each drug are provided in *Supplemental Material* and summarized in **Supplemental\_Fig\_S1**. Next, we performed sequence similarity searches for these genes on the 276 ExPEC draft genomes. Resistance genes and sequence variants found among the 276



genomes are listed in **Supplemental\_Data\_S2**. To investigate the mobility of these genes, we also looked for the presence of known plasmid replicons, listed in **Supplemental\_Data\_S1** by incompatibility group (Carattoli 2011).

#### *Representation of resistance genes in our dataset*

We found a variety of aminoglycoside modification enzymes encoded among the 276 genomes, with a clear predominance of three specific aminoglycoside acetyltransferases: AAC(6')-Ib-cr (34 genomes, 12.7% of the total), AAC(3)-IId (28 genomes, 10.1%) and a novel variant, AAC(3)-Ile.2 (15 genomes, 5.4%), which differs from the AAC(3)-Ile reported by Ho et al. (Ho et al. 2010) only by three amino acid substitutions: F14L, Q275H and K276E. The distribution of these three aminoglycoside acetyltransferases in our dataset is shown in **Fig. 1**, panels 5, 7 and 10.

There were also 19 genes encoding the ANT(2'')-I adenylyltransferase or variants thereof (**Supplemental\_Data\_S2**), and only one encoding the 16S rRNA methyltransferase, *RmtE*. The genome bearing this methyltransferase displayed resistance to both gentamicin and tobramycin and did not contain any other aminoglycoside resistance genes within our consideration, suggesting a causal link to gentamicin and tobramycin resistance. Another genome contained a nonsynonymous *rpsL* mutation (T39I), but it was unclear whether the *rpsL* T39I mutation contributed to gentamicin and tobramycin resistance because AAC(3)-IId was also encoded in this genome. In addition, we found genes corresponding to two aminoglycoside-modifying enzymes that confer resistance to aminoglycosides not included in this study.

There were 117 TEM, 37 OXA, and 48 CTX-M  $\beta$ -lactamases encoded among the 276 genomes. Of the six different types of plasmid-encoded AmpC  $\beta$ -lactamases that we checked, only one was found: the CMY-2  $\beta$ -lactamase (5 genomes, **Supplemental\_Data\_S2**). Note that SHV-1  $\beta$



-lactamases were not represented in any of the genomes. We did not search for carbapenemases because carbapenem resistance was almost completely absent from our clinical sample (<1% prevalence).

We also searched for polymorphisms in the  $\beta$ -lactamases. Among the 117 TEM sequences, there were only three containing ESBL mutations, two with the G238S substitution and one with the R164S mutation. Among the 37 OXA sequences, we discovered a single OXA-1 variant bearing an M102I substitution, which is in the active site and has not been previously reported. CTX-M  $\beta$ -lactamase sequences, on the other hand, exhibited much broader diversity, including CTX-M-15 (37 genomes), CTX-M-14 (6 genomes), CTX-M-27 (3 genomes) and CTX-M-65 and CTX-M-104 (1 genome each). The distribution of CTX-M-15 and of CTX-M-14-like enzymes is shown in **Fig. 1**, panels 6 and 9, respectively.

The profile of  $\beta$ -lactamases represented in our population is consistent with evidence of CTX-M displacement of other ESBL enzymes worldwide (Rodriguez-Villalobos et al. 2011), with CTX-M-14 and CTX-M-15 being the predominant ones (Peirano and Pitout 2010; Hiroi et al. 2012; Doi et al. 2013). Of note, CTX-M-27 and CTX-M-15 were found together in one of our genomes. The coexistence of different CTX-M enzymes in the same sample has been reported, and has already given rise to new CTX-M variants by recombination (Sun et al. 2010; Tian et al. 2014).

Fluoroquinolone resistance generally maps to point mutations in two type-II topoisomerases, although resistance can also be plasmid-mediated (Hopkins et al. 2005; Redgrave et al. 2014). Of the 276 genomes, 47.6% had a *gyrA* mutation (translating to S83L), which is a key resistance-conferring mutation in *E. coli* and several other Gram negatives (Redgrave et al. 2014; Johnning et al. 2015). S83L was frequently accompanied by a mutation in *parC*, S80I/R, which in

combination with GyrA\_S83L produces high levels of fluoroquinolone resistance (Marcusson et al. 2009). The frequent co-occurrence of these two, synergistic mutations is well established (Johnning et al. 2015) and can be clearly seen in **Fig. 1**, panels 2 and 3. By contrast to *gyrA* and *parC* mutants, which were ubiquitous among fluoroquinolone-resistant samples, we only found one plasmid-mediated quinolone resistance gene.

#### *Correlation of genetic markers with drug resistance*

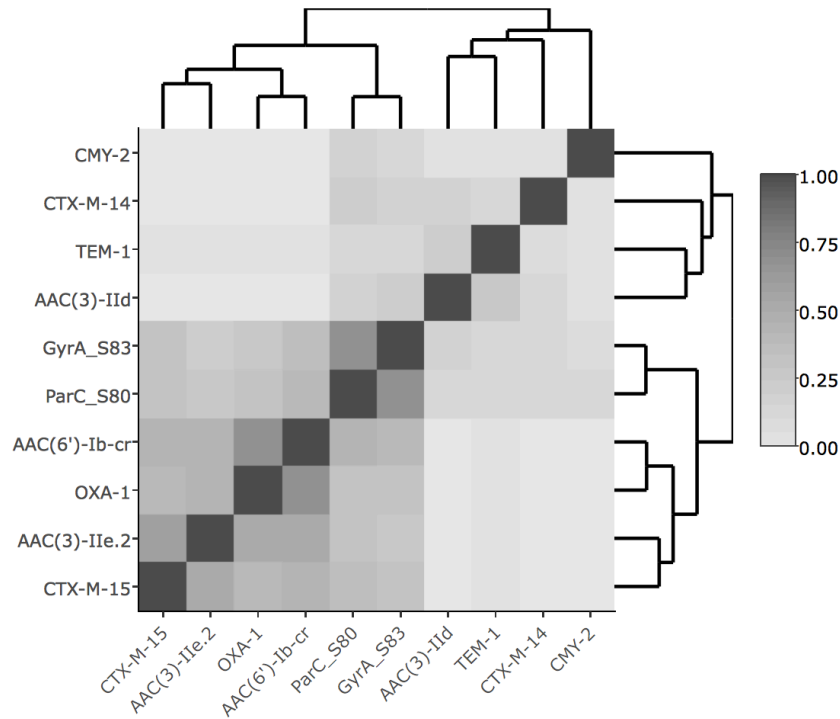
In order to infer the possible selections driving the fixation of resistance genes in our clinical populations, we performed a linear regression analysis (**Supplemental\_Fig\_S2**), which can identify genes that contribute to prediction of resistance for a given drug, and provide a coefficient reflecting the relative contribution of each individual gene to the predictive model. Note that the coefficients for individual genes cannot be interpreted in isolation, as some of these genes are genetically linked, *i.e.* frequently co-occur in samples (**Fig. 1**).

In the case of aminoglycosides (**Supplemental\_Fig\_S2, 2b**), the presence of the two AAC(3)-type acetyltransferase genes correlated preferentially with gentamicin resistance, whereas AAC(6')-Ib-cr genes showed a preferential, although weaker, correlation with tobramycin resistance.  $\beta$ -lactams and their resistance markers displayed more complex relationships (**Supplemental\_Fig\_S2c-h**). CMY-2, CTX-M-14 and CTX-M-15 genes all displayed strong correlations with resistance to the cephalosporins ceftazidime (3<sup>rd</sup> gen), ceftriaxone (3<sup>rd</sup> gen) and cefepime (4<sup>th</sup> gen). Resistance to monobactams (aztreonam, **Supplemental\_Fig\_S2e**), however, correlated only to the CTX-M-15 and CMY-2 genes, suggesting that CTX-M-14 has little activity against monobactams. OXA-1's apparent correlation with resistance to these antibiotics likely arose from linkage disequilibrium between the OXA-1 and CTX-M-15 genes on IncF plasmids (see discussion of complementarity groups below).

No single  $\beta$ -lactamase gene showed a predominant contribution towards prediction of resistance to ampicillin, consistent with the ability of all of these  $\beta$ -lactamases to target ampicillin and with the diversity of  $\beta$ -lactamases encoded in the 276 genomes. For unknown reasons, cefazolin showed a linear regression profile more similar to that of 3<sup>rd</sup> and 4<sup>th</sup>-generation cephalosporins despite being a first-generation cephalosporin, and therefore a substrate for TEM-1, OXA-1, and CMY2 (Jacoby 2009; Gutkind et al. 2013).

#### *Linkage structure of MDR genes*

To get an initial idea of antibiotic resistance gene co-occurrence in the 276 genomes, we generated a heatmap representing the strength of association corresponding to each pair (**Fig. 2**). Aminoglycoside acetyltransferase and ESBL  $\beta$ -lactamase genes clustered to two different groups based on their strengths of association. The first group contained AAC(6')-Ib-cr and AAC(3)-IIe.2, while the second group had AAC(3)-IId. Similarly,  $\beta$ -lactamases also clustered to two groups. CTX-M-15 and OXA-1 clustered to the first group, while TEM-1 and CTX-M-14 family or CMY-2 clustered to the second. Mutations conferring fluoroquinolone resistance (GyrA\_S83L and ParC\_S80I/R), by contrast, were strongly associated with both groups, although showing a slight preference for the first group (**Fig. 2**).



**Figure 2.** Pairwise associations between antibiotic resistance markers estimated based on their co-occurrence in genomes. The heatmap shows the strengths of associations calculated as log odds ratios normalized to between 0.00 (weak) and 1.00 (strong).

We also find that functionally redundant genes within these two groups rarely overlap. The two acetyltransferase genes showing strong correlations with gentamicin resistance (AAC(3)-IId and AAC(3)-IIe.2) do not co-occur in the 33 genomes containing these genes. The same was true when we compared the two acetyltransferase genes that correlated with resistance to tobramycin (AAC(3')-IId and AAC(6')-Ib-cr), which don't co-occur in any of the 40 genomes where they were found. Similarly, we only found one case of co-occurrence between the two groups of ESBL  $\beta$ -lactamase genes seen in our samples: CTX-M-15 and CTX-M-14 family/CMY-2 in a total of 42 relevant genomes. In the case of OXA-1 and TEM-1 genes, the lack of overlap is not so sharp, but still substantial: out of 117 genomes containing TEM and 37 containing OXA, only 10 genomes have both.

In sum, we found that aminoglycoside and  $\beta$ -lactamase genes are genetically linked, although in some cases not very tightly, in two distinct clusters. We also see that when genes are functionally redundant, they rarely overlap. Therefore, the observed combinations of genes can be understood as two mutually exclusive complementarity groups: complementarity group 1 (CG1) and complementarity group 2 (CG2).

*Predictive value of individual genetic markers, alone and in combination*

Next, we explored the predictive value of using antibiotic resistance genes, individually and in combination, as markers for epidemiological monitoring of resistance in the population and also for use in point-of-care diagnostics of antibiotic resistance. If two genes represent functionally redundant solutions for resistance to the same antibiotic, as our gene co-occurrence analysis suggests, each gene by itself should have low sensitivity, while including both genes should dramatically increase the sensitivity of our predictive models. Indeed, inclusion of both AAC(3)-Ile.2 and AAC(3)-IId for prediction of GEN resistance resulted in much greater sensitivity (82.5%) than when each was used as a marker individually (22.2% and 53.3%, respectively). Similarly, AAC(6')-Ib-cr and AAC(3)-IId complement each other for prediction of TOB resistance, producing a sensitivity of 88.9%. For prediction of resistance against 3<sup>rd</sup> and 4<sup>th</sup> generation beta-lactams (CAZ, CRO, FEP), both CTX-M-15 and the CTX-M-14 family of beta-lactamases are needed to achieve high sensitivity (**Supplemental\_Table\_S2, Supplemental\_Table\_S3**).

*Mapping of CG1 and CG2 genes on our cladogram identifies multiple HGT events and allows inference of directionality*

We next mapped our diagnostic CG1 genes encoding OXA-1, AAC(6')-Ib-cr, CTX-M-15, and AAC(3)-IIe.2, and CG2 genes encoding TEM-1 and AAC(3)-IIId on the phylogeny we generated based on SNP variation across the genome (**Fig. 1**). We hypothesized that when the host strains are ordered according to phylogenetic relatedness, the presence of these genes in non-contiguous host strains approximates independent occurrences. However, gene loss cannot be ruled out as an alternative explanation for the discontinuity of representation in the cladogram.

Our analysis identified 16 putative independent occurrences for CG1 (labeled 1 through 16 in **Table 1**) and 22 of them for CG2 (labeled A through V in **Table 2**). The OXA-1 and AAC(6')-Ib-cr genes are found in all 16 putative occurrences, suggesting a strong genetic linkage between these two CG1 genes. Indeed, they are found adjacent to each other and flanked by insertion elements. CTX-M-15 is also almost always found as well. The exception is occurrence #7. Interestingly, in occurrence #2 CTX-M-15 is found in some samples but not all. Taken together, these observations suggest that CTX-M-15 is the third gene to be acquired in this group. AAC(3)-IIe.2 is found only in 9 of the 16 occurrences, and in occurrences with more than one sample (#3 and 9), it is not found in all samples. This suggests that AAC(3)-IIe.2 is the most recently acquired gene into this complementarity group.

**Table 1.** Complementarity group 1.

Putative independent occurrence	MLST	No. of samples	First marker	Second Marker	Third maker	Fourth marker	Other	TEM	ParC S80X	IncFIA
1	131	1	OXA-1	AAC(6')-Ib-cr	CTX-M-15	AAC(3)-IIe.2		TEM-1	1	1
2	131	4	OXA-1	AAC(6')-Ib-cr	(CTX-M-15)			TEM-1	4	4
3	131	11	OXA-1	AAC(6')-Ib-cr	CTX-M-15	(AAC(3)-IIe.2)**			11	11
4	131	1	OXA-1	AAC(6')-Ib-cr				TEM-1	1	1
5	131	1	OXA-1	AAC(6')-Ib-cr	CTX-M-15				1	1
6	131	1	OXA-1	AAC(6')-Ib-cr	CTX-M-15				1	1
7	2521	1	OXA-1	AAC(6')-Ib-cr			CTX-M-14		1	0
8	46	1	OXA-1	AAC(6')-Ib-cr	CTX-M-15	AAC(3)-IIe.2		TEM-1	1	1
9	44(x2) 10	3	OXA-1	AAC(6')-Ib-cr	CTX-M-15	(AAC(3)-IIe.2)			3	3
10	167 (~10)	1	OXA-1	AAC(6')-Ib-cr	CTX-M-15	AAC(3)-IIe.2			1	1
11	617	1	OXA-1	AAC(6')-Ib-cr	CTX-M-15		AAC(3)-IIId		1	1
12	405	1	OXA-1	AAC(6')-Ib-cr	CTX-M-15	AAC(3)-IIe.2			1	0
13	405	1	OXA-1	AAC(6')-Ib-cr	CTX-M-15	AAC(3)-IIe.2		TEM-1	1	1
14	405	1			CTX-M-15	AAC(3)-IIe.2		TEM-1	1	1
15	2659	1	OXA-1	AAC(6')-Ib-cr	CTX-M-15				1	1
16	46	1	OXA-1	AAC(6')-Ib-cr	CTX-M-15	AAC(3)-IIe.2		TEM-1	1	1

**Note:** Markers with incomplete penetrance are shown in parenthesis; putative independent event #9 likely represents two events, because one of the strains involved belongs to a different MLST type; frequency of diagnostic markers by themselves: OXA-1 (2), AAC(6')-Ib-cr (2), CTX-M-15 (10); AAC(2)-II2.e (0).

**Table 2.** Complementarity group 2.

Putative independent occurrence	MLST	No. of samples	First marker	Second marker	Other	ParC- S80X (GyrA-S83L)	IncFA1
A	131	2	TEM-1	AAC(3)-IId		0 (2)	0
B	131	1	TEM-1	AAC(3)-IId		0	0
C	131	1	TEM-1	AAC(3)-IId		1	1
D	131	1	TEM-1	AAC(3)-IId		1	1
E	131	1	TEM-1	AAC(3)-IId		1	1
F	131	1	TEM-1	AAC(3)-IId		1	1
G	131	1	TEM-1	AAC(3)-IId		1	1
H	131	2	TEM-1	AAC(3)-IId		2	2
I	131	1	TEM-1	AAC(3)-IId		1	1
J	162	1	TEM-1		CTX-M-14	1	0
K	5612	1	TEM-1	AAC(3)-IId		1	0
L	10	1	TEM-1	AAC(3)-IId		1	0
M	N/A	1	TEM-1		CTX-M-14 CTX-M-15	0	0
N	69	1	TEM-1	AAC(3)-IId		1	1
O	69	1	TEM-1		CTX-M-14	1	1
P	393	1	TEM-1	AAC(3)-IId		1	1
Q	648	2	TEM-1	AAC(3)-IId	CTX-M-14	2	0
R	354	1	TEM-1	AAC(3)-IId	CTX-M-14	1	1
S	Unknown	1	TEM-1	AAC(3)-IId		(1)	1
T	73	2	TEM-1	AAC(3)-IId		0	2
U	69	1	TEM-1		CTX-M-14	1	1
V	1193	2	TEM-1	AAC(3)-IId*		2	2

**Note:** Frequency of diagnostic markers by themselves within this group: AAC 3 IId (1); CTX-M-14 (4); TEM (74); AA( 3)-IId (0) or CTX-M-14 (0).



CG2 is characterized by the presence of AAC(3)-IId as opposed to other AAC enzymes, by a tight linkage to TEM-1 (100% of occurrences, compared to an average frequency of 44% in our population and in GC1 samples), and by a near complete absence of CTX-M-15.

*IncF plasmids are frequent and carry resistance to oxymino cephalosporins and aminoglycosides*

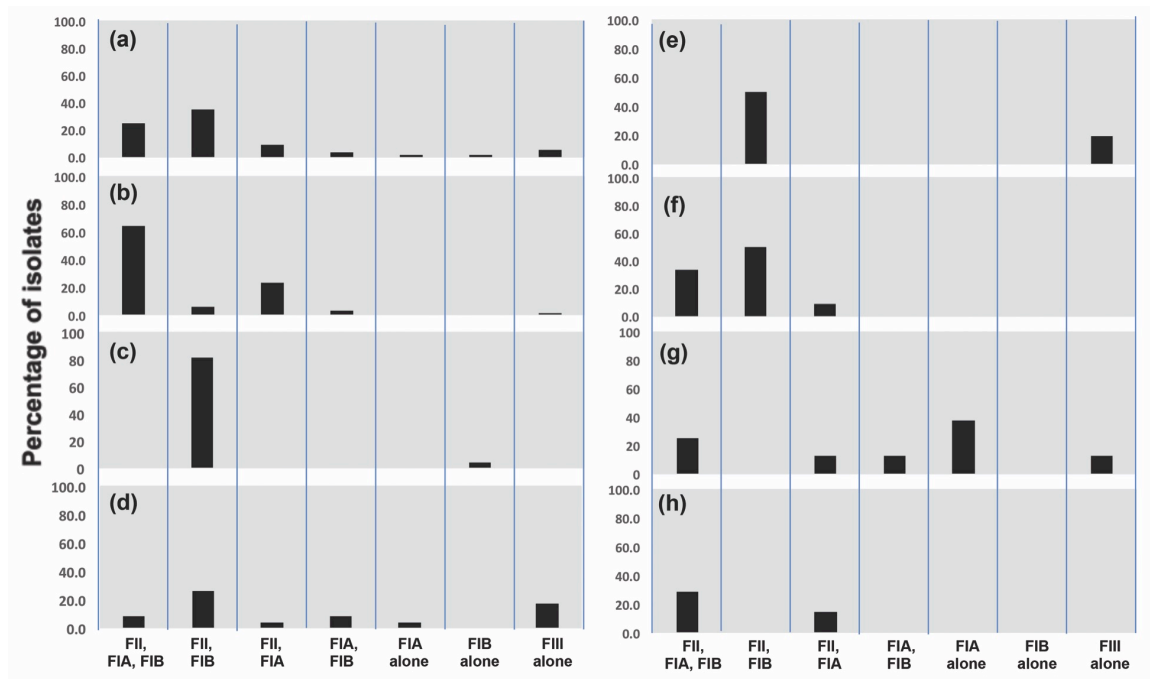
IncF plasmids were very prevalent in our clinical population: 221 out of 276 strains carried an IncF plasmid, as determined by the presence of IncFIA, IncFIB or IncFII origins of replication.

CG1 and CG2 genes appear to be carried by IncF plasmids. In our data, aminoglycoside and ESBL resistance was found almost exclusively in the 221 genomes where IncF replicons were detected (with only 2 exceptions in the 55 remaining genomes). Further, IncF replicons were the only replicons consistently associated with AAC and plasmid-borne  $\beta$ -lactamase genes even though IncF-bearing genomes contained a variety of other replicons: IncI1, IncK, IncP, IncBO, IncAC, IncN, ColE1, IncR, IncY and IncLM (**Supplemental\_Data\_S1**).

The role of IncF plasmids in the acquisition of bla<sub>TEM-1</sub> and bla<sub>OXA-1</sub>, in the subsequent capture of ESBL genes (Woodford et al. 2011; Olesen et al. 2013), and in the dissemination of CTX-M genes (Nicolas-Chanoine et al. 2008; Woodford et al. 2009; Hu et al. 2014) and of CMY (Matsumura et al. 2012; Naseer et al. 2012) by ST131 strains is well-established. Their role in spreading AAC(6')-Ib-cr and CTX-M has been reported (Baudry et al. 2009; Li et al. 2015; Agyekum et al. 2016), and the AAC-(6')-Ib-cr, OXA-1, and CTX-M15 combination has also been previously mapped to IncF plasmids (Li et al. 2015).

IncF plasmid origins of replication are modular (Hu et al. 2014). IncFII seems to be the essential element, present in nearly all plasmids. IncFIB is next in order of frequency (177), followed by

IncFIA (106). **Fig. 3a** shows the average distribution of IncF origins of replication for our entire sample population. IncFIA is most frequently associated with both IncFII and IncFIB, whereas IncFIB is most frequently associated with IncFII alone, without IncFIA. This pattern of occurrence is consistent with a previous report (Hu et al. 2014) and suggests an asymmetrical dependence of IncFIA on IncFIB or a differential frequency of loss between the two subgroups.



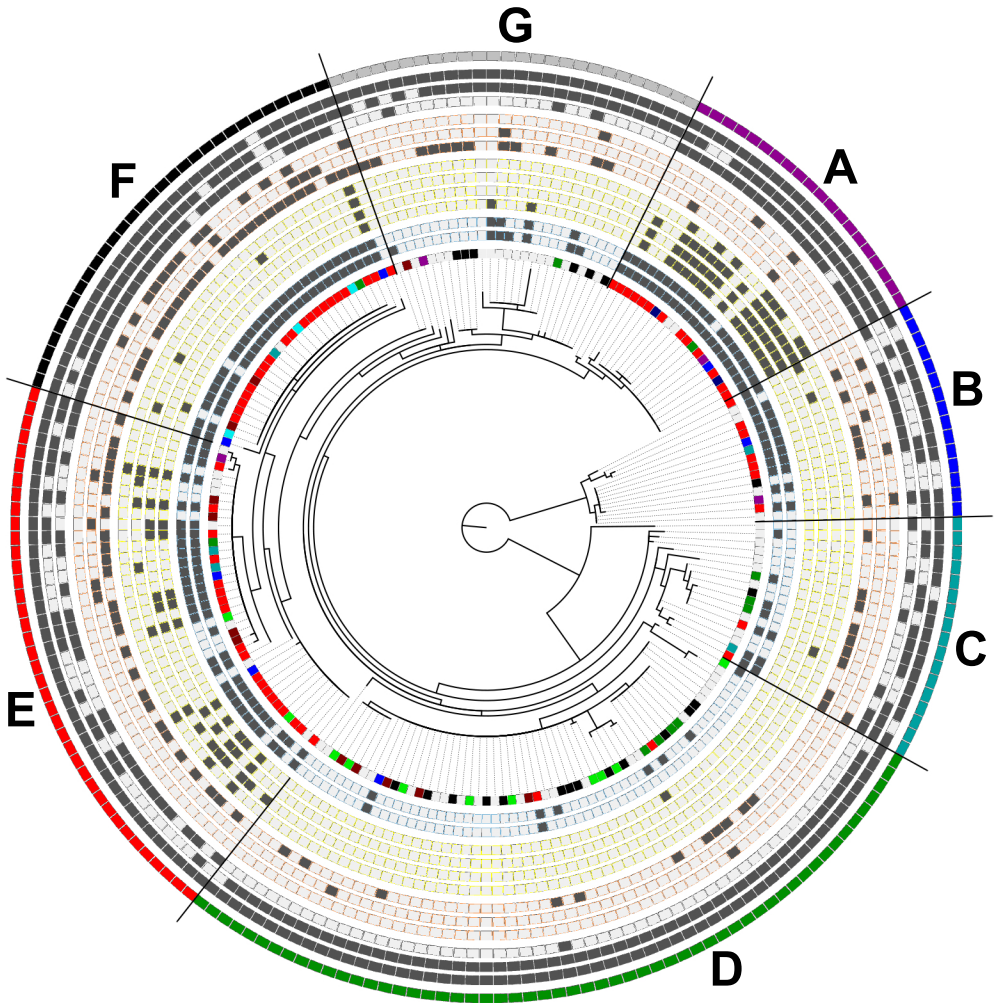
**Figure 3.** Distribution of IncF plasmid replicon type combinations across several prominent MLST groups. Proportion of isolates (y-axis) containing each combination of IncF replicon types is shown across the 276 ExPEC isolates, **(a)** as a reference, or in the samples belonging to the following MLST types: **(b)** ST131, **(c)** ST95, **(d)** ST173, **(e)** ST127, **(f)** ST69, **(g)** ST393 and **(h)** ST10.

**Fig. 3b-h** shows the average distribution of IncF origins of replication for the seven MLST types that are represented with at least 6 samples: ST131 (**panel b**), ST95 (**panel c**), ST173 (**panel d**),

ST127 (**panel e**), ST69 (**panel f**), ST393 (**panel g**) and ST10 (**panel h**). The representation of IncFIA is strikingly variable depending on the MLST: from not represented at all (ST95, and ST127), to greatly overrepresented (ST131, ST10). IncFIA overrepresentation can be due to an increased representation of IncFIA alone (ST393) or of IncFIA in combination with IncFII and/or IncFIB (ST131, ST10).

#### *IncF plasmids form distinct clades*

Next, we investigated the phylogenetic structure of the IncF plasmids present in our population. Given that plasmids are subject to frequent recombination and HGT, we focused our analysis on a single protein, the Rpt-A1 replicase, which is found in a region of the IncF plasmid that appears to be relatively stable (Villa et al. 2010; Shin et al. 2012). To further minimize noise from HGT and gene duplications, we limited our phylogeny to 204 samples carrying full-length (not truncated) Rpt-A1 sequences found as single copy genes (of 221 total samples carrying IncF plasmids). This analysis indicated the presence of seven distinct classes of IncF plasmids (**Fig. 4**). To reinforce our conclusions from our phylogenetic analysis of a single IncF plasmid protein (Rpt-A1), we performed Gaussian Mixture Modeling with Expectation Maximization to generate a second grouping of plasmid sequences based on the presence or absence of multiple IncF-specific sequence elements including various replication origins, replicase genes and relaxase genes. This analysis identified nine clusters (**Supplemental\_Fig\_S3**).



**Figure 4.** Maximum-likelihood phylogeny of 203 IncF-specific replicases. Rpt-A1. Genomic assemblies encoding multiple Rpt-A1 of canonical length, but with diverged sequences, were not included in the phylogeny in order to minimize noise from HGT and gene duplications. The Circos diagram surrounding the phylogeny shows for each phylogenetic taxa represented by one unambiguous Rpt-A1 sequence, its MLST classification, presence of various resistance markers and plasmid replicons. Circular panels represent the following, from the innermost to outermost position. **Panel 1**, MLST classification: ST131 (red), ST44, (navy), ST10 (purple), ST405 (blue), ST69 (green), ST393 (teal), ST648 (cyan), ST73 (maroon), ST95 (black), ST127 (lime). **Panels 2** and **3** (outlined in blue) fluoroquinolone resistance mutations: GyrA-S83X and ParC-S80I, respectively. **Panels 4-8** (outlined in yellow), complementarity group 1: genes for OXA-1, AAC(6')-Ib-cr, CTX-M-15, and AAC(3)-II2.e, respectively. **Panels 9-11** (outlined in orange), complementarity group 2: genes for TEM-1, CTX-M-9-family and AAC(3)-II2.d, respectively.

**Panels 12-14** (outlined in grey), IncF replicons: IncFIA, IncFIB, IncFII, respectively. **Panel 15**, phylogenetic clades assigned manually upon careful inspection of Rpt-A1 phylogeny clade structure.

*Detection of IncF horizontal and vertical transfer*

Given that we previously established that MLSTs closely approximate phylogenetic relatedness in our population (**Fig. 1**), we investigated whether the seven classes of plasmid defined above were restricted in their host strain distribution, focusing on MLSTs that are well represented in our population (with more than five samples). The number of different plasmid classes found in each of these MLSTs is shown in **Supplemental\_Fig\_S4a**. This number ranged from 2 to 6, which is very high considering that some of these MLSTs were represented by fewer than 10 genomes. The distribution of plasmid groupings obtained from our Gaussian Mixture Modeling analysis across MLSTs also suggested extensive conjugation among different plasmid groups (**Supplemental\_Fig\_S5**). The most parsimonious explanation for the presence of different classes of IncF plasmids in strains belonging to the same type is HGT by conjugation.

The presence of plasmids belonging to the same class in samples that are contiguous in the cladogram and that belong to the same MLST is suggestive of clonal expansion, because in this case vertical transmission of the plasmid from a common ancestor is the most parsimonious explanation. This analysis points again to the four frequent MLSTs identified earlier as likely representing recent clonal expansion, namely ST131, ST95, ST73, and ST127. In addition, this analysis found two additional, less frequent MLSTs (3 samples each) that we would not have otherwise been able to detect: ST648, and ST421.

*CG1 and CG2 are carried by distinct sets of IncF plasmids*

We mapped the occurrences of the genes defined in CG1 and CG2 to our Rpt-A1-based phylogeny of IncF plasmids (**Fig. 4**) with the assumption that these genes are largely carried in IncF plasmids. The gene-to-isolate correspondence is provided in **Supplemental\_Table\_S4**. We found that each CG is carried by distinct sets of IncF plasmids.

CG1 is largely found in classes A and E. In class A, the linkage between genes encoding OXA-1, AAC(6')-Ib-cr and CTX-M-15 is stronger than in class E (100% vs 50%). Class E also has a higher representation of CTX-M-14-like genes (4 samples, compared to 1 for Class A). Clade F has one instance of the complete set of CG1 genes including one encoding AAC(3)-Ile.2, suggesting that despite the inconsistent linkage between the genes that define CG1, they can be moved together.

For CG1, there are two additional associations: with fluoroquinolone resistance mutations, and with IncFIA. ParC-S80I/R is present in 79.6% of the samples, but in 100% of the samples carrying CG1 genes. Likewise, while IncFIA represents 72% of the total combining IncF plasmid classes A and E, it is present in 93% of the samples carrying CG1 genes. These two associations are specific for CG1: for CG2, ParC-S80I/R was found in 74% of the samples (77% expected combining classes B and F) and h IncF1A was found in 74% of the samples (67% expected).

CG2 corresponds to clades B and F in our plasmid phylogeny. Clade F has the same frequency of CTX-M-14 and of CTX-M-15, though.

Finally, we looked to see whether the different classes of IncF plasmids were differentially distributed in clonally expanding strains. **Supplemental\_Fig\_S5b** shows the breakdown of classes represented in our frequent MLSTs (black columns), compared to the distribution found in

rare MLSTs (white columns). All seven classes have a good representation in both groups of strains, again consistent with ubiquitous HGT. We found that the classes associated with GC1 and CG2 drug resistance (A, B, E and F) are overrepresented in clonally expanding strains, whereas classes C and G (with a weaker association with drug resistance) are more frequent in rare strains. This seems to corroborate the idea that high risk clones help spread resistance.

## Discussion

Here we describe the presence of two sets of genes in clinical samples from two hospitals on the West Coast of the US: the first set (CG1) includes four genes encoding drug-inactivating enzymes: OXA-1, AAC(6')-Ib-cr, CTX-M-15 and AAC(3)-IIe.2. The second set (CG2) includes genes for TEM-1, AAC(3)-IId and CTX-M-14-like enzymes.

Using linear regression analysis, we established strong correlations between the presence of these genes and resistance to some aminoglycosides and oxyimino-cephalosporins (**Supplemental Fig S2**), and using these genes, constructed strong predictive models for these drugs (**Supplemental Table S2** and **Supplemental Table S3**). Tobramycin and ampicillin were exceptions, suggesting the presence of additional mechanisms of resistance to these drugs that we have not been able to identify.

The repeated identification of sequencing assemblies corresponding to a few genes, with conserved boundaries across different contexts (multiresistance regions or MRRs) has been described in the past (Partridge 2011). The genes belonging to our two CGs likely comprise MMRs, although we have not verified their location in the sequence relative to each other (with the exception of AAC(6')-Ib-cr and OXA-1 which occur in close proximity, and bounded by conserved insertion elements). These CG genes do co-occur much more frequently than would be expected by chance (**Tables 1, 2** and **Fig. 1**) and parts of the two CGs that we describe here have

already been reported as MRRs: the co-occurrence of AAC-6'-Ib-cr with CTX-M-15 (Gibreel et al. 2012; Banerjee et al. 2013), and the combination of AAC-6'-Ib-cr, OXA-1, and CTX-M-15 (Li et al. 2015).

Here we report three novel findings regarding the linkage between these resistance genes: (1) that CTX-M-14-like genes are found in a separate CG relative to CTX-M-15; (2) that a new AAC(3) aminoglycoside acetyltransferase closely related to AAC(3)-IIe (AAC(3)-IIe.2) is in the process of being acquired by CG1; and (3) that each CG is linked to a different  $\beta$ -lactamase gene: OXA-1 (CG1) and TEM-1 (CG2), although TEM-1 is ubiquitous.

We also show that these previously reported gene associations are very prevalent and explain the bulk of gentamicin and ESBL resistance cases in our two U.S. West Coast hospitals. Remarkably, this includes not only ST131 types but also many others (**Tables 1, 2**).

ST131 is unique in that it is a carrier for genes belonging to both CGs with similar frequency: 10% of ST131 isolates carry CG1 genes and 14% carry CG2 genes. However, for the most part, CG1 and CG2 have different host types. Frequent CG1 carriers include ST44, ST46, and ST405; others include ST617, ST2521, and ST2659. By contrast, ST69 is the most frequent carrier for CG2 genes (25% of 12 isolates); other CG2 hosts include: ST73, ST354, ST393, ST1193 and ST5612. These findings extend previous reports indicating an association of specific MLSTs with CTX-M enzyme dissemination: ST131 and ST405 in the case of CTX-M-15 (Woodford et al. 2011; Canton et al. 2012; Shin et al. 2012), and ST393 in the case of CTX-M-14 (Canton et al. 2012). The multiplicity of MLSTs seen is consistent with the widespread mobilization of IncF plasmids across types shown in **Fig. 4**.



We find that when CG1 and CG2 genes are functionally redundant, they almost never overlap in the same isolate. This pattern suggests that the gene set fixed in a given population is largely the result of contingent evolution; once a strain acquires one MRR, an alternative MRR no longer provides a selective advantage. This is the reason why we propose calling these linked gene sets “complementarity groups”. The exceptions are OXA-1 and TEM-1, which in this case appear to follow a mutually-exclusive distribution pattern because of their linkage to AAC(6’)-Ib-cr (CG1) and AAC(3)-IIId (CG2), respectively. Thus, in this case mutual exclusion is indirect, and therefore weaker.

However, the linkage between the genes in each complementarity group is dynamic, which allows us to infer the order in which the resistance markers are being acquired following the principle of genetic capitalism originally postulated by (Baquero 2004), **Fig. 1** and **Tables 1, 2**.

In the case of CG1, the initial mutations appear to be mutations in GyrA and ParC conferring fluoroquinolone resistance. This had been previously reported in the context of ST131 (Naseer and Sundsfjord 2011), but here we show that this phenomenon applies more broadly to strains harboring CG1 genes (**Table 1**, second column). The OXA-1 + AAC(6’)-Ib-cr combination is the next one to appear; the two genes are adjacent and flanked by insertion sequences, which explains their tight linkage. The next gene to be captured is CTX-M-15, almost certainly as a result of treatment with oxymino cephalosporins or with aztreonam. Finally, the last gene, which is very much in the process of capture into CG1 is AAC(3)-IIe.2, likely as a result of gentamicin selection.

CG2 appears to have arisen by the capture of AAC(3)-IIId, linked to TEM-1. Subclass II AAC(3) acetyltransferases are common and confer resistance to both tobramycin and gentamicin (Shaw et al. 1993), consistent with our linear regression analysis, which suggests that this gene provides

resistance to both tobramycin and gentamicin (**Supplemental\_Fig\_S2**). The next gene, which is in the process of capture is CTX-M-14 in the broad sense (also including other enzymes of the CTX-M-9 cluster).

The apparent connection between fluoroquinolone resistance (which is chromosomal) and the IncF-borne CG1 aminoglycoside and ESBL resistance genes is not obvious and intriguing. It could simply be the result of the timing of selection, with fluoroquinolones being used as first line of defense and ESBLs as an alternative when fluoroquinolones fail, but this does not explain why this pattern is not seen for the CG2 group. Alternatively, AAC(6')-Ib-cr, which confers fluoroquinolone resistance, may be selected under fluoroquinolone selective pressure as an enhancer of the resistance provided by chromosomal mutations in GyrA and ParC. This would explain, why a separate AAC (AAC(3)-IIe.2) gene is selected later on to provide gentamicin resistance. This would also explain why in strains harboring CG2 genes, where no specific linkage to fluoroquinolone resistance can be detected above a high background of resistance, a single gene that provides resistance to both gentamycin and tobramycin was captured.

We identified a few sequence types that are undergoing clonal expansion and that therefore constitute potential high-risk clones (Woodford et al. 2011): ST131, ST95, ST73, ST127, and ST69. This conclusion is based on representation (all have at least 12 samples), on a short phylogenetic distance between samples belonging to the same type, and on evidence of vertical transmission of CG genes (since this only happens following acquisition of this putative MRR and is therefore a relatively recent evolutionary event). Further, we observed that ST131 strains were much more abundant in the DHMMC samples (which were obtained more recently), representing 67.4% of the total, compared to 16.5% in the UW study (which was 2-3 years older), consistent with the idea that ST131 strains continue to expand in the clinical population. We confirmed that this apparent overrepresentation of ST131 strains in the DHMMC set cannot be

attributed to its different antibiotic resistance profile, by performing Monte Carlo sampling distributions matching the proportions of ceftazidime-, ciprofloxacin- or gentamycin-resistant samples among the 43 DHMMC assemblies (see *Methods*, *Supplemental Materials* and **Supplemental\_Fig\_S6**).

The ST131 type is known to have undergone a dramatic expansion since they were first reported in all three continents in 2008 (Simner et al. 2011; Colpan et al. 2013). It is interesting to note that the three high-risk clones with the highest representation in our study (ST131, ST73 and ST95) all belong to the phylogroup B2 (Le Gall et al. 2007), suggesting that this expansion is not simply driven by selection for drug resistance (many other STs also have these genes) but by some ecological/virulence advantage. There is strong evidence that this is the case for ST131 strains, whose expansion is at least partially driven by the acquisition of FimH, a mannose-binding type 1 fimbrial adhesin, which appears to facilitate uroepithelial colonization by *E. coli* (Johnson et al. 2013). On the other hand, we did find IncF plasmid classes associated with CG1 and CG2 drug resistance (A, B, E and F) are overrepresented in clonally-expanding strains, supporting a role for these strains in the spread of antibiotic resistance. The fact that only a subset of plasmids belonging to these classes carry CG1 or CG2 genes suggests that they are facilitating the spread of resistance but not necessarily driving it.

IncF plasmid replicons are modular and complex, frequently including multiple typing subgroups (Hu et al. 2014). Similar to other studies (Villa et al. 2010), we find that IncFII is the essential one, frequently associated with IncFIB, which in turn can be associated with IncFIA (**Fig. 1, 4**). Here we show that IncF replicon modules can be highly MSLT-specific (**Fig. 3**).

We also show an association of IncFIA with the plasmid classes bearing CG1 and CG2 genes (frequency of IncFIA representation range of 60-95%, compared to 2-35% for the other

plasmids); in classes A and E, there was an additional association between CG1 resistance genes and IncFIA: 93% of strains carrying CG1 genes also carried IncFIA origins of replication, compared to 72% of strains carrying class A and E plasmids. These observations suggest a preferential role of IncFIA origins of replication in the mobilization of antibiotic resistance.

Our Rpt-A1 phylogeny also found 13 instances of likely co-occurrence of IncF plasmids, based on the identification of two Rpt-A1 copies in the same sample (**Supplemental Table S5**). In all but one case, the two copies belonged to a different plasmid class, indicating that these do not represent recent duplication events. Interestingly, in all but two cases, one of the plasmids belonged to class B, the one class that is clearly more phylogenetically different from the other six (**Fig. 4**). This observation suggests that class B plasmids may be frequently compatible with other IncF classes. The observation that IncF plasmid co-occurrence appears to depend on the phylogenetic distance of Rpt-A1 suggests that compatibility is not a process primarily driven by the IncF subgroup complement. Instead, the modularity of IncF plasmid replicons may represent an adaptation to the host rather. This would be consistent with the observation of IncF signatures for individual MLST types reported here.

## Concluding remarks

Here we describe a binary gene distribution of  $\beta$ -lactamase and aminoglycoside resistance genes in the clinical population of two West-Coast hospitals. This observation suggests an evolutionary convergence between different evolutionary solutions, producing ubiquitous functional redundancy. The discovery of specific complementarity groups and the functional redundancy they represent has significant implications for monitoring and controlling the spread of drug resistance. For example, functional redundancy greatly complicates the rational design of antibiotic therapeutic regimens aimed at maintaining the population sensitive (Mira et al. 2015).

We were able to build models with high predictive accuracies (90-95%) for most drugs using combinations of only 2-3 genes. Sequence-based predictions offer three critical advantages over clinical microbiology assays (1) reproducibility, *i.e.* by using a standardized analytic pipeline, results can be reproduced which is important for clinical-related usage, (2) traceable historical data, which helps monitor the spread of resistance, and (3) contextual information, which allows estimation on the likelihood of resistance development to additional antibiotics [reviewed in (Didelot et al. 2012; Burnham et al. 2017)]. However, the extent to which the markers that we have identified can be generalized to other clinical populations is unclear.

Accurate prediction of resistance phenotype at point of care would help delay the spread of resistance, minimizing misuse of antibiotics, which currently increases both the costs of treating infection and accelerates the spread of resistance. On the other hand, evidence of pervasive HGT and recombination indicates a great potential for the acquisition and spread of new resistance genes that would need to be incorporated to our predictive models if these are to remain accurate.

## **Acknowledgements**

We would like to acknowledge Dr . Steve Salipante’s help processing and interpreting the UW sequences, Dr. Gerardo Cortés-Cortés for useful comments on the manuscript. This work was funded by CITRIS Seed Funding proposal 2015-324 to TL, MC and MB, and by NIH/NIAID award 1R41AI122740-01A1 to PC, MC and MB.

## **Disclosure declaration**

The authors don’t have any conflict of interest to disclose.

## **Methods**

### **Sample and data collection**

Clinical samples were collected from patients with respiratory, blood (wound), or urinary tract infections at Dignity Health Mercy Medical Center (DHMMC) in Merced, California, between June 2013 and August 2015. The isolates were identified as ESBL-positive using an automated rapid detection system for pathogen identification and antibiotic sensitivity, Vitek 2 Version 06.01. Following identification, the samples were also tested for susceptibility against 16 antibiotics using broth micro-dilution minimum inhibitory concentration (MIC) testing. The isolates were categorized according to their susceptibility: Resistant (R), Intermediate (I), or Susceptible (S), based on the MIC Interpretation Guideline – CLSI M100-S26 (2015). The 16 antibiotics included 1 penicillin: Ampicillin, 2 penicillin and inhibitor combinations: Ampicillin/Sulbactam, Piperacillin/Tazobactam, 4 cephalosporins: Cefazolin, Ceftazidime, Ceftriaxone, Cefepime, 2 carbapenems: Ertapenem, Imipenem, 3 aminoglycosides: Amikacin,

Gentamycin, Tobramycin, 2 fluoroquinolones: Ciprofloxacin, Levofloxacin, and Nitrofurantoin and Trimetoprim/Sulfamethoxazole.

Additionally, we obtained 342 ExPEC genome assemblies from a previous study conducted at the University of Washington (UW), downloaded from Genbank (Benson et al. 2016) with accessions in the range JSFQ00000000–JSST00000000.

### **Sequencing, quality control and assembly of ExPEC genomes**

Genomic DNA was extracted from each sample using the ZR-96 Quick-gDNA Kit from Zymo Research. Whole genome sequencing including TruSeq DNA library preparation was conducted at the University of California, Davis Genome Center using Illumina's MiSeq and HiSeq technologies. We obtained 24 MiSeq (2x250 bp) and 24 HiSeq (2x250 bp) paired-end sequencing libraries corresponding to our selected samples. Prior to assembly, Illumina sequencing adapters and low quality bases were trimmed from the sequencing reads using Trimmomatic v0.36 (Bolger et al. 2014); trimmed reads shorter than 36 bp were discarded. Library quality was also verified using FastQC v0.11.5. *De novo* paired-end assembly was conducted for each MiSeq and HiSeq library using SPAdes v3.5.0 (Bankevich et al. 2012) with read error correction by BWA-spades. The spades.py wrapper script was used to select an appropriate *k*-mer size for optimized assembly of each genome.

The resulting assemblies were scanned for contamination and misassembly; contigs shorter than 300 bp or with low coverage were discarded. Estimated genome sizes (median: 5.3 Mb) based on the remaining contigs were generally within the range of known ExPEC genomes. The assemblies had 29x and 327x median coverage for MiSeq and HiSeq libraries, respectively. The median N50 across 22 MiSeq and 23 HiSeq library assemblies was 198 kb (median L50: 9) and 201 kb

(median L50: 8.5). Two HiSeq libraries appeared to contain significant contamination or proportions of misassembled contigs based on estimated genome sizes (8.2 Mb and 9.0 Mb for samples 328 and 357, respectively); these two assemblies were removed from downstream analyses. We further assessed each DHMMC and UW assemblies for completeness based on the presence of 143 protein-coding genes considered to be essential for normal growth of *E. coli* MG1655 [Set A in Supplementary Table 2 from (Gerdes et al. 2003)]. To achieve consistent assembly qualities across the DHMMC and UW datasets, we removed assemblies containing fewer than 126 full-length essential genes.

### **Phylogeny and typing of strains**

We conducted multi-locus sequence typing (MLST) of the 276 genome assemblies based on 15 housekeeping loci required for categorizing *E. coli* strains according to the Achtman (7 loci) and Pasteur (8 loci) MLST schemes.

To estimate phylogenetic relationships among the 276 assemblies, we constructed a neighbor-joining tree based on all possible pair-wise comparisons. For each pairing, an evolutionary distance based on single nucleotide polymorphisms (SNPs) was estimated using the nesoni package [ref]. SNPs for individual assemblies were detected by aligning all sequencing reads comprising an assembly against an *E. coli* reference genome, EC958, which belongs to the ST131 (Forde et al. 2014). Note that a phylogeny generated in this manner may be subject to noise arising from SNPs found among horizontally transferred elements (e.g., genes in plasmids and genomic islands). However, we believe such noise should have little influence on the shape of the phylogeny since the number of SNPs found in vertically transmitted regions should greatly outnumber the SNPs found in horizontally transmitted regions. We confirmed the general accuracy of our SNP-based phylogeny by comparison against a maximum-likelihood phylogeny generated using the concatenated protein sequences of 126 essential *E. coli* genes.



A second maximum-likelihood phylogeny based on Rpt-A1 protein sequence homology was generated to approximate the evolutionary relatedness of IncF plasmids. To minimize noise from HGT and gene duplications, we first removed assemblies putatively encoding multiple, different Rpt-A1 proteins of canonical length (285 amino acids) *i.e.* assemblies with multiple Rpt-A1 were included only if there was 100% sequence agreement. Maximum-likelihood phylogenies were generated using RAxML v8.2.10 (Stamatakis 2014) ; bootstrapping was performed on the CIPRES compute cluster (Miller et al. 2015). Multiple alignments for these trees were generated using MAFFT v7.294b (Kato and Standley 2013); gaps were removed using TrimAl (Capella-Gutierrez et al. 2009).

### **Gene model prediction and annotation**

Gene models were predicted using Prodigal v2.6.2 (Hyatt et al. 2010) in metagenomic mode. Known antibiotic resistance-conferring genes and “mutations” were annotated using NCBI BLASTx and BLASTp (Altschul et al. 1990) based on amino acid residue-level differences between selected “wild-type” genes and full-length hits (greater than or equal to 99% query coverage and greater than 90 percent identity). Selected “wild-type” genes included known resistance genes encoding  $\beta$ -lactamases (TEM-1, SHV-1, OXA-1, CTX-M-14, CTX-M-15), aminoglycoside-modifying enzymes (acetyltransferases [AAC], adenytransferases [ANT], phosphoryltransferases [APH]), 30S ribosomal protein subunit S12, 16S rRNA methyltransferases (ArmA, RmtA, RmtB, RmtC, RmtD, RmtE), DNA gyrase subunits (GyrA and GyrB), DNA topoisomerase IV subunits (ParC and ParE), and plasmid-mediated quinolone resistance proteins (QnrA, QnrB, QnrB2, QnrC, QnrD, QnrS). Amino acid sequences and Uniprot (UniProt 2015) accessions for wild-type genes are provided in **Supplemental\_Data\_S2**. Plasmid replication origins and associated incompatibility groups were annotated based on BLASTn hits (e-value <  $10^{-50}$ ) against INC-DB (Lanza et al. 2014), a plasmid origin nucleotide sequence

database.

### **Predictive values and co-occurrence of known genes**

The predictive values of individual genes and combinations of genes were assessed based on Pearson correlation and several other measures including sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV). The relative importance of each marker for resistance prediction was estimated as regression coefficients obtained from a generalized logistic regression prediction model, fitted using the glm package in R. Reported coefficients were the average result of 1000 bootstrap iterations. Marker co-occurrence associations were assessed using adjusted p-values and log odds ratios calculated using Fisher's exact tests.

### **References**

- Agyekum A, Fajardo-Lubian A, Ansong D, Partridge SR, Agbenyega T, Iredell JR. 2016. blaCTX-M-15 carried by IncF-type plasmids is the dominant ESBL gene in *Escherichia coli* and *Klebsiella pneumoniae* at a hospital in Ghana. *Diagn Microbiol Infect Dis* **84**: 328-333.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol* **215**: 403-410.
- Banerjee R, Johnson JR. 2013. *Escherichia coli* ST131: variations on a theme of clonal expansion. *Enferm Infec Microbiol Clin* **31**: 355-356.
- Banerjee R, Strahilevitz J, Johnson JR, Nagwekar PP, Schora DM, Shevrin I, Du H, Peterson LR, Robicsek A. 2013. Predictors and molecular epidemiology of community-onset extended-spectrum beta-lactamase-producing *Escherichia coli* infection in a Midwestern community. *Infect Control Hosp Epidemiol* **34**: 947-953.
- Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Pribelski AD et al. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* **19**: 455-477.
- Banuelos-Vazquez LA, Torres Tejerizo G, Brom S. 2017. Regulation of conjugative transfer of plasmids and integrative conjugative elements. *Plasmid* **91**: 82-89.

- Baquero F. 2004. From pieces to patterns: evolutionary engineering in bacterial pathogens. *Nat Rev Microbiol* **2**: 510-518.
- Baudry PJ, Nichol K, DeCorby M, Lagace-Wiens P, Olivier E, Boyd D, Mulvey MR, Hoban DJ, Zhanel GG. 2009. Mechanisms of resistance and mobility among multidrug-resistant CTX-M-producing *Escherichia coli* from Canadian intensive care units: the 1st report of QepA in North America. *Diagn Microbiol Infect Dis* **63**: 319-326.
- Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW. 2016. GenBank. *Nucleic Acids Res* doi:10.1093/nar/gkw1070.
- Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**: 2114-2120.
- Briales A, Rodriguez-Martinez JM, Velasco C, de Alba PD, Rodriguez-Bano J, Martinez-Martinez L, Pascual A. 2012. Prevalence of plasmid-mediated quinolone resistance determinants qnr and aac(6')-Ib-cr in *Escherichia coli* and *Klebsiella pneumoniae* producing extended-spectrum beta-lactamases in Spain. *Int J Antimicrob Agents* **39**: 431-434.
- Brolund A, Sandegren L. 2016. Characterization of ESBL disseminating plasmids. *Infect Dis (Lond)* **48**: 18-25.
- Burnham CD, Leeds J, Nordmann P, O'Grady J, Patel J. 2017. Diagnosing antimicrobial resistance. *Nat Rev Microbiol* **15**: 697-703.
- Cag Y, Caskurlu H, Fan Y, Cao B, Vahaboglu H. 2016. Resistance mechanisms. *Ann Transl Med* **4**: 326.
- Cagnacci S, Gualco L, Debbia E, Schito GC, Marchese A. 2008. European emergence of ciprofloxacin-resistant *Escherichia coli* clonal groups O25:H4-ST 131 and O15:K52:H1 causing community-acquired uncomplicated cystitis. *Journal of clinical microbiology* **46**: 2605-2612.
- Canton R, Coque TM, Baquero F. 2003. Multi-resistant Gram-negative bacilli: from epidemics to endemics. *Curr Opin Infect Dis* **16**: 315-325.
- Canton R, Gonzalez-Alba JM, Galan JC. 2012. CTX-M Enzymes: Origin and Diffusion. *Front Microbiol* **3**: 110.
- Canton R, Ruiz-Garbajosa P. 2011. Co-resistance: an opportunity for the bacteria and resistance genes. *Curr Opin Pharmacol* **11**: 477-485.
- Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**: 1972-1973.
- Carattoli A. 2011. Plasmids in Gram negatives: molecular typing of resistance plasmids. *Int J Med Microbiol* **301**: 654-658.
- Carattoli A. 2013. Plasmids and the spread of resistance. *Int J Med Microbiol* **303**: 298-304.
- Cerceo E, Deitelzweig SB, Sherman BM, Amin AN. 2016. Multidrug-Resistant Gram-Negative Bacterial Infections in the Hospital Setting: Overview, Implications for Clinical Practice, and Emerging Treatment Options. *Microb Drug Resist* **22**: 412-431.
- Coll F, McNerney R, Preston MD, Guerra-Assuncao JA, Warry A, Hill-Cawthorne G, Mallard K, Nair M, Miranda A, Alves A et al. 2015. Rapid determination of

- anti-tuberculosis drug resistance from whole-genome sequences. *Genome Med* **7**: 51.
- Colpan A, Johnston B, Porter S, Clabots C, Anway R, Thao L, Kuskowski MA, Tchesnokova V, Sokurenko EV, Johnson JR et al. 2013. Escherichia coli sequence type 131 (ST131) subclone H30 as an emergent multidrug-resistant pathogen among US veterans. *Clin Infect Dis* **57**: 1256-1265.
- Crofts TS, Gasparrini AJ, Dantas G. 2017. Next-generation approaches to understand and combat the antibiotic resistome. *Nat Rev Microbiol* **15**: 422-434.
- Didelot X, Bowden R, Wilson DJ, Peto TEA, Crook DW. 2012. Transforming clinical microbiology with bacterial genome sequencing. *Nat Rev Genet* **13**: 601-612.
- Doi Y, Park YS, Rivera JI, Adams-Haduch JM, Hingwe A, Sordillo EM, Lewis JS, 2nd, Howard WJ, Johnson LE, Polsky B et al. 2013. Community-associated extended-spectrum beta-lactamase-producing Escherichia coli infection in the United States. *Clin Infect Dis* **56**: 641-648.
- Doi Y, Wachino JI, Arakawa Y. 2016. Aminoglycoside Resistance: The Emergence of Acquired 16S Ribosomal RNA Methyltransferases. *Infect Dis Clin North Am* **30**: 523-537.
- Fitzpatrick MA, Ozer EA, Hauser AR. 2016. Utility of Whole-Genome Sequencing in Characterizing Acinetobacter Epidemiology and Analyzing Hospital Outbreaks. *Journal of clinical microbiology* **54**: 593-612.
- Forde BM, Ben Zakour NL, Stanton-Cook M, Phan MD, Totsika M, Peters KM, Chan KG, Schembri MA, Upton M, Beatson SA. 2014. 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* **9**: e104400.
- Gerdes SY, Scholle MD, Campbell JW, Balazsi G, Ravasz E, Daugherty MD, Somera AL, Kyrpidis NC, Anderson I, Gelfand MS et al. 2003. Experimental determination and system level analysis of essential genes in Escherichia coli MG1655. *J Bacteriol* **185**: 5673-5684.
- Ghafourian S, Sadeghifard N, Soheili S, Sekawi Z. 2015. Extended Spectrum Beta-lactamases: Definition, Classification and Epidemiology. *Curr Issues Mol Biol* **17**: 11-21.
- Gibreel TM, Dodgson AR, Cheesbrough J, Fox AJ, Bolton FJ, Upton M. 2012. Population structure, virulence potential and antibiotic susceptibility of uropathogenic Escherichia coli from Northwest England. *J Antimicrob Chemother* **67**: 346-356.
- Gordon NC, Price JR, Cole K, Everitt R, Morgan M, Finney J, Kearns AM, Pichon B, Young B, Wilson DJ et al. 2014. Prediction of Staphylococcus aureus antimicrobial resistance by whole-genome sequencing. *Journal of clinical microbiology* **52**: 1182-1191.
- Guedon G, Libante V, Coluzzi C, Payot S, Leblond-Bourget N. 2017. The Obscure World of Integrative and Mobilizable Elements, Highly Widespread Elements that Pirate Bacterial Conjugative Systems. *Genes (Basel)* **8**.
- Gutkind GO, Di Conza J, Power P, Radice M. 2013. beta-lactamase-mediated resistance: a biochemical, epidemiological and genetic overview. *Curr Pharm Des* **19**: 164-208.

- Hiroi M, Yamazaki F, Harada T, Takahashi N, Iida N, Noda Y, Yagi M, Nishio T, Kanda T, Kawamori F et al. 2012. Prevalence of extended-spectrum beta-lactamase-producing *Escherichia coli* and *Klebsiella pneumoniae* in food-producing animals. *J Vet Med Sci* **74**: 189-195.
- Ho PL, Wong RC, Lo SW, Chow KH, Wong SS, Que TL. 2010. Genetic identity of aminoglycoside-resistance genes in *Escherichia coli* isolates from human and animal sources. *J Med Microbiol* **59**: 702-707.
- Hopkins KL, Davies RH, Threlfall EJ. 2005. Mechanisms of quinolone resistance in *Escherichia coli* and *Salmonella*: recent developments. *Int J Antimicrob Agents* **25**: 358-373.
- Hu F, O'Hara JA, Rivera JI, Doi Y. 2014. Molecular features of community-associated extended-spectrum-beta-lactamase-producing *Escherichia coli* strains in the United States. *Antimicrob Agents Chemother* **58**: 6953-6957.
- Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**: 119.
- Ilangovan A, Connery S, Waksman G. 2015. Structural biology of the Gram-negative bacterial conjugation systems. *Trends Microbiol* **23**: 301-310.
- Jacoby GA. 2009. AmpC beta-lactamases. *Clin Microbiol Rev* **22**: 161-182, Table of Contents.
- Jana S, Deb JK. 2006. Molecular understanding of aminoglycoside action and resistance. *Appl Microbiol Biotechnol* **70**: 140-150.
- Jeon JH, Lee JH, Lee JJ, Park KS, Karim AM, Lee CR, Jeong BC, Lee SH. 2015. Structural basis for carbapenem-hydrolyzing mechanisms of carbapenemases conferring antibiotic resistance. *Int J Mol Sci* **16**: 9654-9692.
- Johnning A, Kristiansson E, Fick J, Weijdegard B, Larsson DG. 2015. Resistance Mutations in *gyrA* and *parC* are Common in *Escherichia* Communities of both Fluoroquinolone-Polluted and Uncontaminated Aquatic Environments. *Front Microbiol* **6**: 1355.
- Johnson JR, Tchesnokova V, Johnston B, Clabots C, Roberts PL, Billig M, Riddell K, Rogers P, Qin X, Butler-Wu S et al. 2013. Abrupt emergence of a single dominant multidrug-resistant strain of *Escherichia coli*. *J Infect Dis* **207**: 919-928.
- Johnson TJ, Wannemuehler YM, Johnson SJ, Logue CM, White DG, Doetkott C, Nolan LK. 2007. Plasmid replicon typing of commensal and pathogenic *Escherichia coli* isolates. *Appl Environ Microbiol* **73**: 1976-1983.
- Kang CI, Cha MK, Kim SH, Ko KS, Wi YM, Chung DR, Peck KR, Lee NY, Song JH. 2013. Clinical and molecular epidemiology of community-onset bacteremia caused by extended-spectrum beta-lactamase-producing *Escherichia coli* over a 6-year period. *J Korean Med Sci* **28**: 998-1004.
- Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* **30**: 772-780.
- Lanza VF, de Toro M, Garcillan-Barcia MP, Mora A, Blanco J, Coque TM, de la Cruz F. 2014. Plasmid flux in *Escherichia coli* ST131 sublineages, analyzed by plasmid constellation network (PLACNET), a new method for plasmid reconstruction from whole genome sequences. *PLoS genetics* **10**: e1004766.

- Le Gall T, Clermont O, Gouriou S, Picard B, Nassif X, Denamur E, Tenaillon O. 2007. Extraintestinal virulence is a coincidental by-product of commensalism in B2 phylogenetic group *Escherichia coli* strains. *Mol Biol Evol* **24**: 2373-2384.
- Leinberger DM, Grimm V, Rubtsova M, Weile J, Schroppel K, Wichelhaus TA, Knabbe C, Schmid RD, Bachmann TT. 2010. Integrated detection of extended-spectrum-beta-lactam resistance by DNA microarray-based genotyping of TEM, SHV, and CTX-M genes. *Journal of clinical microbiology* **48**: 460-471.
- Li JJ, Spychala CN, Hu F, Sheng JF, Doi Y. 2015. Complete nucleotide sequences of bla(CTX-M)-harboring IncF plasmids from community-associated *Escherichia coli* strains in the United States. *Antimicrob Agents Chemother* **59**: 3002-3007.
- Li XZ, Mehrotra M, Ghimire S, Adewoye L. 2007. beta-Lactam resistance and beta-lactamases in bacteria of animal origin. *Vet Microbiol* **121**: 197-214.
- Luo T, Yang C, Peng Y, Lu L, Sun G, Wu J, Jin X, Hong J, Li F, Mei J et al. 2014. Whole-genome sequencing to detect recent transmission of *Mycobacterium tuberculosis* in settings with a high burden of tuberculosis. *Tuberculosis (Edinb)* **94**: 434-440.
- Magnet S, Blanchard JS. 2005. Molecular insights into aminoglycoside action and resistance. *Chem Rev* **105**: 477-498.
- Maiden MC, Bygraves JA, Feil E, Morelli G, Russell JE, Urwin R, Zhang Q, Zhou J, Zurth K, Caugant DA et al. 1998. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc Natl Acad Sci U S A* **95**: 3140-3145.
- Mairi A, Pantel A, Sotto A, Lavigne JP, Touati A. 2017. OXA-48-like carbapenemases producing Enterobacteriaceae in different niches. *Eur J Clin Microbiol Infect Dis* doi:10.1007/s10096-017-3112-7.
- Marcusson LL, Frimodt-Moller N, Hughes D. 2009. Interplay in the selection of fluoroquinolone resistance and bacterial fitness. *PLoS Pathog* **5**: e1000541.
- Martinez-Martinez L, Gonzalez-Lopez JJ. 2014. Carbapenemases in Enterobacteriaceae: types and molecular epidemiology. *Enferm Infecc Microbiol Clin* **32 Suppl 4**: 4-9.
- Matsumura Y, Yamamoto M, Nagao M, Hotta G, Matsushima A, Ito Y, Takakura S, Ichiyama S, Kyoto-Shiga Clinical Microbiology Study G. 2012. Emergence and spread of B2-ST131-O25b, B2-ST131-O16 and D-ST405 clonal groups among extended-spectrum-beta-lactamase-producing *Escherichia coli* in Japan. *J Antimicrob Chemother* **67**: 2612-2620.
- Miller MA, Schwartz T, Pickett BE, He S, Klem EB, Scheuermann RH, Passarotti M, Kaufman S, O'Leary MA. 2015. A RESTful API for Access to Phylogenetic Tools via the CIPRES Science Gateway. *Evol Bioinform Online* **11**: 43-48.
- Mira PM, Crona K, Greene D, Meza JC, Sturmfels B, Barlow M. 2015. Rational design of antibiotic treatment plans: a treatment strategy for managing evolution and reversing resistance. *PLoS One* **10**: e0122283.
- Naseer U, Olsson-Liljequist BE, Woodford N, Dhanji H, Canton R, Sundsfjord A, Lindstedt BA. 2012. Multi-locus variable number of tandem repeat analysis for rapid and accurate typing of virulent multidrug resistant *Escherichia coli* clones. *PLoS One* **7**: e41232.



- Naseer U, Sundsfjord A. 2011. The CTX-M conundrum: dissemination of plasmids and Escherichia coli clones. *Microb Drug Resist* **17**: 83-97.
- Nicolas-Chanoine MH, Bertrand X, Madec JY. 2014. Escherichia coli ST131, an intriguing clonal group. *Clin Microbiol Rev* **27**: 543-574.
- Nicolas-Chanoine MH, Blanco J, Leflon-Guibout V, Demarty R, Alonso MP, Canica MM, Park YJ, Lavigne JP, Pitout J, Johnson JR. 2008. Intercontinental emergence of Escherichia coli clone O25:H4-ST131 producing CTX-M-15. *J Antimicrob Chemother* **61**: 273-281.
- Novick RP. 1987. Plasmid incompatibility. *Microbiol Rev* **51**: 381-395.
- Olesen B, Hansen DS, Nilsson F, Frimodt-Moller J, Leihof RF, Struve C, Scheutz F, Johnston B, Krogfelt KA, Johnson JR. 2013. Prevalence and characteristics of the epidemic multiresistant Escherichia coli ST131 clonal group among extended-spectrum beta-lactamase-producing E. coli isolates in Copenhagen, Denmark. *Journal of clinical microbiology* **51**: 1779-1785.
- Partridge SR. 2011. Analysis of antibiotic resistance regions in Gram-negative bacteria. *FEMS Microbiol Rev* **35**: 820-855.
- Partridge SR, Zong Z, Iredell JR. 2011. Recombination in IS26 and Tn2 in the evolution of multiresistance regions carrying blaCTX-M-15 on conjugative IncF plasmids from Escherichia coli. *Antimicrob Agents Chemother* **55**: 4971-4978.
- Peirano G, Pitout JD. 2010. Molecular epidemiology of Escherichia coli producing CTX-M beta-lactamases: the worldwide emergence of clone ST131 O25:H4. *Int J Antimicrob Agents* **35**: 316-321.
- Peirano G, van der Bij AK, Gregson DB, Pitout JD. 2012. Molecular epidemiology over an 11-year period (2000 to 2010) of extended-spectrum beta-lactamase-producing Escherichia coli causing bacteremia in a centralized Canadian region. *Journal of clinical microbiology* **50**: 294-299.
- Price LB, Johnson JR, Aziz M, Clabots C, Johnston B, Tchesnokova V, Nordstrom L, Billig M, Chattopadhyay S, Stegger M et al. 2013. The epidemic of extended-spectrum-beta-lactamase-producing Escherichia coli ST131 is driven by a single highly pathogenic subclone, H30-Rx. *MBio* **4**: e00377-00313.
- Ramirez MS, Tolmasky ME. 2010. Aminoglycoside modifying enzymes. *Drug Resist Updat* **13**: 151-171.
- Redgrave LS, Sutton SB, Webber MA, Piddock LJ. 2014. Fluoroquinolone resistance: mechanisms, impact on bacteria, and role in evolutionary success. *Trends Microbiol* **22**: 438-445.
- Robicsek A, Jacoby GA, Hooper DC. 2006. The worldwide emergence of plasmid-mediated quinolone resistance. *Lancet Infect Dis* **6**: 629-640.
- Rodriguez-Bano J, Gutierrez-Gutierrez B, Machuca I, Pascual A. 2018. Treatment of Infections Caused by Extended-Spectrum-Beta-Lactamase-, AmpC-, and Carbapenemase-Producing Enterobacteriaceae. *Clin Microbiol Rev* **31**.
- Rodriguez-Martinez JM, Machuca J, Cano ME, Calvo J, Martinez-Martinez L, Pascual A. 2016. Plasmid-mediated quinolone resistance: Two decades on. *Drug Resist Updat* **29**: 13-29.
- Rodriguez-Villalobos H, Bogaerts P, Berhin C, Bauraing C, Deplano A, Montesinos I, de Mendonca R, Jans B, Glupczynski Y. 2011. Trends in production of

- extended-spectrum beta-lactamases among Enterobacteriaceae of clinical interest: results of a nationwide survey in Belgian hospitals. *J Antimicrob Chemother* **66**: 37-47.
- Rubtsova MY, Ulyashova MM, Bachmann TT, Schmid RD, Egorov AM. 2010. Multiparametric determination of genes and their point mutations for identification of beta-lactamases. *Biochemistry (Mosc)* **75**: 1628-1649.
- Salipante SJ, Roach DJ, Kitzman JO, Snyder MW, Stackhouse B, Butler-Wu SM, Lee C, Cookson BT, Shendure J. 2015. Large-scale genomic sequencing of extraintestinal pathogenic Escherichia coli strains. *Genome research* **25**: 119-128.
- Sauvage E, Terrak M. 2016. Glycosyltransferases and Transpeptidases/Penicillin-Binding Proteins: Valuable Targets for New Antibacterials. *Antibiotics (Basel)* **5**.
- Shakil S, Khan R, Zarrilli R, Khan AU. 2008. Aminoglycosides versus bacteria--a description of the action, resistance mechanism, and nosocomial battleground. *J Biomed Sci* **15**: 5-14.
- Shaw KJ, Rather PN, Hare RS, Miller GH. 1993. Molecular genetics of aminoglycoside resistance genes and familial relationships of the aminoglycoside-modifying enzymes. *Microbiol Rev* **57**: 138-163.
- Shin J, Choi MJ, Ko KS. 2012. Replicon sequence typing of IncF plasmids and the genetic environments of blaCTX-M-15 indicate multiple acquisitions of blaCTX-M-15 in Escherichia coli and Klebsiella pneumoniae isolates from South Korea. *J Antimicrob Chemother* **67**: 1853-1857.
- Simner PJ, Zhanel GG, Pitout J, Taitel F, McCracken M, Mulvey MR, Lagace-Wiens PR, Adam HJ, Hoban DJ, Canadian Antimicrobial Resistance A. 2011. Prevalence and characterization of extended-spectrum beta-lactamase- and AmpC beta-lactamase-producing Escherichia coli: results of the CANWARD 2007-2009 study. *Diagn Microbiol Infect Dis* **69**: 326-334.
- Snitkin ES, Zelazny AM, Thomas PJ, Stock F, Group NCSP, Henderson DK, Palmore TN, Segre JA. 2012. Tracking a hospital outbreak of carbapenem-resistant Klebsiella pneumoniae with whole-genome sequencing. *Sci Transl Med* **4**: 148ra116.
- Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**: 1312-1313.
- Sun Y, Zeng Z, Chen S, Ma J, He L, Liu Y, Deng Y, Lei T, Zhao J, Liu JH. 2010. High prevalence of bla(CTX-M) extended-spectrum beta-lactamase genes in Escherichia coli isolates from pets and emergence of CTX-M-64 in China. *Clin Microbiol Infect* **16**: 1475-1481.
- Tian GB, Huang YM, Fang ZL, Qing Y, Zhang XF, Huang X. 2014. CTX-M-137, a hybrid of CTX-M-14-like and CTX-M-15-like beta-lactamases identified in an Escherichia coli clinical isolate. *J Antimicrob Chemother* **69**: 2081-2085.
- UniProt C. 2015. UniProt: a hub for protein information. *Nucleic Acids Res* **43**: D204-212.
- Ventola CL. 2015. The antibiotic resistance crisis: part 1: causes and threats. *P T* **40**: 277-283.



- Villa L, Garcia-Fernandez A, Fortini D, Carattoli A. 2010. Replicon sequence typing of IncF plasmids carrying virulence and resistance determinants. *J Antimicrob Chemother* **65**: 2518-2529.
- Walker TM, Kohl TA, Omar SV, Hedge J, Del Ojo Elias C, Bradley P, Iqbal Z, Feuerriegel S, Niehaus KE, Wilson DJ et al. 2015. Whole-genome sequencing for prediction of Mycobacterium tuberculosis drug susceptibility and resistance: a retrospective cohort study. *Lancet Infect Dis* **15**: 1193-1202.
- Woodford N, Carattoli A, Karisik E, Underwood A, Ellington MJ, Livermore DM. 2009. 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* **53**: 4472-4482.
- Woodford N, Turton JF, Livermore DM. 2011. Multiresistant Gram-negative bacteria: the role of high-risk clones in the dissemination of antibiotic resistance. *FEMS Microbiol Rev* **35**: 736-755.
- Yanat B, Rodriguez-Martinez JM, Touati A. 2017. Plasmid-mediated quinolone resistance in Enterobacteriaceae: a systematic review with a focus on Mediterranean countries. *Eur J Clin Microbiol Infect Dis* **36**: 421-435.
- Yano H, Uemura M, Endo S, Kanamori H, Inomata S, Kakuta R, Ichimura S, Ogawa M, Shimojima M, Ishibashi N et al. 2013. Molecular characteristics of extended-spectrum beta-lactamases in clinical isolates from Escherichia coli at a Japanese tertiary hospital. *PLoS One* **8**: e64359.
- Yelin I, Kishony R. 2018. Antibiotic Resistance. *Cell* **172**: 1136-1136 e1131.
- Zapun A, Contreras-Martel C, Vernet T. 2008. Penicillin-binding proteins and beta-lactam resistance. *FEMS Microbiol Rev* **32**: 361-385.
- Zowawi HM, Harris PN, Roberts MJ, Tambyah PA, Schembri MA, Pezzani MD, Williamson DA, Paterson DL. 2015. The emerging threat of multidrug-resistant Gram-negative bacteria in urology. *Nat Rev Urol* **12**: 570-584.