Dynamic Periodicity of Extended-Spectrum Cephalosporin Resistant Bloodstream *Escherichia coli* Infections Driven by Non-CC131 Strains

Running Title: Molecular Epidemiology of ESC-R *E. coli* BSIs in Hematological Patients

William C Shropshire[a], Benjamin Strope[b], Selvalakshmi Selvaraj Anand[b], Jordan Bremer[a], Patrick McDaneld[c*], Micah M Bhatti[d], Anthony R Flores[e], Awdhesh Kalia[b], Samuel A Shelburne[a,f#]

[a]Department of Infectious Diseases and Infection Control, The University of Texas MD Anderson Cancer Center, Houston, TX, USA

[b]Program in Diagnostic Genetics and Genomics, MD Anderson Cancer Center School of Health Professions, Houston, TX, USA

[c]Division of Pharmacy, The University of Texas MD Anderson Cancer Center, Houston, TX, USA

[d]Department of Laboratory Medicine, The University of Texas MD Anderson Cancer Center, Houston, TX, USA

[e]Division of Infectious Diseases, Department of Pediatrics, McGovern Medical School, University of Texas Health Science Center, Houston, TX, USA

[f]Department of Genomic Medicine, The University of Texas MD Anderson Cancer Center, Houston, TX, USA

*Patrick McDaneld is currently an employee of bioMérieux Diagnostics but was not an employee at the time of his contributions to the study.

#Address for correspondence: Samuel A Shelburne, MD, PhD, 1901 East Rd., South Campus Research Building, 4SCR5.1042, Houston, TX 77054; phone: (713) 792-3629, e-mail: sshelburne@mdanderson.org

Abstract Word Count: 248

Importance Word Count: 148

Main Text Word Count: 5814

ABSTRACT

Extended-spectrum cephalosporin resistant (ESC-R) *Escherichia coli* (ESC-R-*Ec*) is an urgent public health threat with clonal complex (CC) 131, phylogroup B2 strains being particularly concerning as the dominant cause of ESC-R-*Ec* infections. To address the paucity of recent ESC-R-*Ec* molecular epidemiology data in the United States (US), we used whole genome sequencing (WGS) to fully characterize a large cohort of invasive ESC-R-*Ec* at a tertiary care cancer center in Houston, Texas collected from 2016-2020. During the study timeframe, there were 1154 index *E. coli* bloodstream infections (BSIs) of which 389 (33.7%) were ESC-R. Using time series analyses, we identified a temporal dynamic of ESC-R *E. coli* BSIs (*Ec*-BSIs), distinct from ESC-susceptible *Ec*-BSIs, with cases peaking in the last 6 months of the calendar year. WGS of 297 ESC-R *Ec*-BSI strains revealed that while CC131 strains accounted for ~45% of total infections, the proportion of CC131 strains remained stable across the time-period, and infection peaks were driven by genetically diverse, non-CC131 isolates. Bla_{CTX-M} variants accounted for most β-lactamases conferring the ESC-R phenotype (89%; 220/248 index ESC-R *Ec*-BSIs), and amplification of bla_{CTX-M} genes was widely detected in ESC-R *Ec*-BSI strains, particularly in carbapenem non-susceptible strains and in strains causing recurrent BSIs. $Bla_{CTX-M-55}$ was significantly enriched within phylogroup A strains, and we identified $bla_{CTX-M-55}$ plasmid-to-chromosome transmission occurring across non-B2 strains. Our data provide important information regarding the current molecular epidemiology of invasive ESC-R *E. coli* and provide novel insights into the genetic basis of observed temporal variability for these clinically important pathogens.

IMPORTANCE

Given that *E. coli* is the leading cause of ESC-R *Enterobacterales* infection, we sought to assess the current molecular epidemiology of ESC-R-*Ec* using a WGS based analysis of many BSIs over a several year period. We identified a clear temporal dynamic of ESC-R-*Ec* infections,

which has also recently been identified in other geographical regions such as Israel. Our WGS data allowed us to visualize the stable nature of CC131 over the study period and demonstrate that non-CC131 strains drove the infection peaks. Additionally, we provide the first widespread assessment of β-lactamase gene copy number in ESC-R-*Ec* infections and delineate mechanisms by which such amplifications are achieved in a diverse array of ESC-R-*Ec* strains. These data suggest that even for a tertiary care center, serious ESC-R-*Ec* infections are driven by a diverse array of strains and impacted by environmental factors suggesting that community-based monitoring could inform novel preventative measures.

INTRODUCTION

Extended-spectrum β-lactamase (ESBL) producing *Enterobacterales* (ESBL-E) were ranked as a serious antimicrobial resistance (AMR) threat by the Centers for Disease Control and Prevention in 2019 (1). A recent US study of multi-drug resistant (MDR) infections in hospitalized patients indicated a 53.3% increased incidence in estimated ESBL-E cases from 2012 to 2017 (2). Other sources of United States (US) data have indicated similar trends, such as a 138% increase in resistance to 3rd generation cephalosporins amongst invasive *E. coli* strains from 2009 to 2016 (3, 4). Furthermore, multiple population-based studies indicate that ESBL-E bloodstream infections may have a higher likelihood of adverse outcomes compared to non-ESBL-E bloodstream infections (5-7).

The key impact of ESBL enzymes is 3rd generation cephalosporin resistance (*i.e.,* extended-spectrum cephalosporin resistance; ESC-R). Although ESBL enzymes account for the vast majority of the associated ESC-R phenotype, there are other mechanisms in *Enterobacterales*, such as plasmid-borne AmpC enzymes (8). The use of ESC-R has been increasingly employed in place of ESBL positivity (9), and it is the term we will use for the remainder of this manuscript with the understanding that the two terms are not completely interchangeable. Amongst ESC-R *Enterobacterales*, some 50% or greater of infections are due to *Escherichia coli* (10, 11). ESC-R *E. coli* (ESC-R-*Ec*) in the 1980s and 1990s primarily harbored ESBL variants of TEM and SHV enzymes (12); however, this changed in the early 2000s with the increasing spread of CTX-M enzymes (13). The rise of CTX-M enzyme producing *E. coli* coincided with a marked increase in sequence type (ST) 131 strains (14). In particular, fluoroquinolone resistant sub-populations of ST131 (*i.e.,* ST131-*H*30 subclade C) strongly associated with $bla_{\text{CTX-M-15}}$ and $bla_{\text{CTX-M-27}}$ carriage, have been shown to be the predominant cause of ESC-R-*Ec* infections in the US (15). However, whether the increasing number of ESC-R-*Ec* infections in the US is driven by further expansion of ST131 strains is not well understood.

Application of WGS to large cohorts of temporally diverse *E. coli*, including ESBL producing isolates, from local and nationwide surveys has revealed common as well as unique molecular epidemiology features (16-22). Consistently, *E. coli* strains causing human infections are pauci-clonal with domination of a small number of STs relative to the overall population (16-18, 23). Many analyses indicate ST131 rose to become the leading *E. coli* ST isolated from humans by 2010 (19, 24). However, studies in England and Calgary have found that ST131 has remained at a relatively steady proportion of overall *E. coli* infections over the past five to ten years (16, 19, 22) whereas investigations in France, Spain, and Norway have identified a continued increase in emergent ST131 subclades (17, 20, 21). Geographical variation in relative ST131 proportion changes may reflect differential ST131 cladal adaptation within ecological niches, with advantageous traits failing to fixate in populations due to negative frequency-dependent selection, reflected by the eventual stability of ST131 within the broader *E. coli* population (25).

A comprehensive US study found that ST131 accounted for 50% of all ESC-R-*Ec* strains in 2012 with an estimated increase of 50% from 2007 to 2011 in the percentage of ST131 *E. coli* (26). Given the lack of recent longitudinal data analyzing ESC-R-*Ec* in the US, we sought to gain insight into the molecular epidemiology of ESC-R-*Ec* by analyzing bloodstream infection causing strains at our institution from 2016-2020.

RESULTS

Observed peaks in ESC-R-*Ec* BS occurred temporally within our sampling frame.

We identified total *Escherichia coli* bloodstream infections (*Ec*-BSI) at our institution that occurred from March 1st, 2016, to December 31st, 2020. There were 1370 total *Ec*-BSIs with 84.2% (1154/1370) being index *Ec*-BSI cases and 15.8% (216/1370) being recurrent (see methods for definitions). The overall ESC-R-*Ec* proportion was 33.7% (389/1154) among index *Ec*-BSI cases. The proportion of ESC-R-*Ec* infections that were recurrent (121/510, 23.7%) was significantly higher compared to recurrent extended-spectrum cephalosporin susceptible *E. coli* (ESC-S-*Ec*) cases (95/860, 11%, χ^2 *P*-value < 0.001).

When analyzing the temporal trends of ESC-R-*Ec* within this 4+ year timeframe, the frequency of index ESC-R-*Ec* BSI cases varied yearly from a low of 70 in 2020 to a peak of 99 in 2019 (Table S1). Similarly, the percent of ESC-R-*Ec* strains causing total *Ec*-BSI ranged from a low of 31% in 2020 to a peak of 37% in 2019 (Table S1). Given that the changes in absolute yearly ESC-R-*Ec* might be reflective of the variance in the total population seen at our institution, we assessed the number of monthly admissions. We found a statistically significant increase (linear regression one-sample t-test *P*-value<0.001) in the average number of monthly admissions between the start of the study of period and the beginning of the COVID pandemic (March 2020) before a marked decrease for the remainder of 2020.

Using the number of monthly admissions as a surrogate of the hospital population to calculate prevalence rates, we found no statistically significant increase in ESC-R-*Ec* rates over the course of the study (linear regression one-sample t-test *P*-value=0.20). Notably, when calculating *Ec*-BSI prevalence rates stratified by ESC status, we identified an oscillating pattern of total *Ec*-BSI prevalence rates corresponding to peaks and valleys (dotted grey loess curve; **Figure 1**) that appeared to be driven primarily by ESC-R-*Ec* (solid red loess curve; **Figure 1**)

rather than ESC-S-*Ec* (solid blue curve, **Figure 1**). Stratifying the sampling frame based on these observed increased infection peaks (**Figure 1; Table 1**), ESC-R-*Ec* prevalence rates were greatest during window 2 (1.89 BSI cases/1000 admissions), window 4 (2.09 BSI cases/1000 admissions), and window 6 (2.24BSI cases/1000 admissions) versus an average of 1.29 BSI cases/1000 admissions during non-peak periods ($P < 0.001$ for monthly ESC-R-*Ec* rates during peak vs. non-peak windows by Student's t-test).

We performed time series analyses and found that while both ESC-R-*Ec* and ESC-S-*Ec* had non-significant monotonic trends, there was evidence of a non-monotonic trend for ESC-R-*Ec* that occurred up to the start of the COVID-19 pandemic in March 2020 (WAVK test-statistic =2.7; *P*-value = 0.016), which was not evident in ESC-S-*Ec* (WAVK test-statistic =0.2; *P*-value = 0.877). To determine whether ESC-R-*Ec* and ESC-S-*Ec* prevalence rates were distinct, we compared pairwise month-to-month prevalence and found no statistically significant correlation (R=0.14, Pearson *P*-value = 0.31; Fig. S1A). Given the periodic trends noted in **Figure 1,** we assessed whether variation in *Ec*-BSIs corresponded to the time of year by dividing the index infections into the 1st or 2nd half of the year and found a significantly higher average of ESC-R-*Ec* (*P*-value = 0.003) but not ESC-S-*Ec* (*P*-value =0.09) occurring in the last 6 months relative to the initial 6 months of each year (Fig. S1B). Taken together, these data trends suggest temporal dynamics in ESC-R-*Ec* independent of ESC-S-*Ec* strains with a potential seasonal component.

Peaks in ESC-R *Ec*-BSIs are not due to CC131 strains.

From the 389 index and 121 recurrent ESC-R-*Ec* strains identified during the study period, we generated assemblies for 297 ESC-R-*Ec* isolates including 64% (248/389) of index and 41% (49/121) of recurrent ESC-R-*Ec* strains. The quality of assemblies is shown in Table S2. In total there were 7 phylogroups detected across the full sampling frame (**Table 2**). The majority of index ESC-R-*Ec* belonged to phylogroup B2 (50.8%; 126/248), followed by A (19.8%; 49/248),

D (12.1%; 30/248), F (7.7%; 19/248), B1 (6.5%; 16/248), C (2.8%; 7/248), and G (0.4%; 1/248). We identified a similar distribution for recurrent ESC-R-*Ec* strains with B2 (38.8%; 19/49) as the most common phylogroup followed by A (34.7%; 17/49), D (12.2%; 6/49), B1 (8.2%; 4/49), C (4.1%; 2/49), and F (2.0%; 1/49). In terms of sequence types (STs) (**Table 2**), ST131 was the most common for both index (n = 102, 41%) and recurrent strains (n = 17, 35%). A further breakdown of each of the most common STs detected by phylogroup can be found in **Table 2**.

To investigate the molecular epidemiology underlying the temporal dynamics of our ESC-R-*Ec* population, we analyzed isolates characterized by phylogroup across infection windows (**Figure 2**). **Figure 2A-B** provides an illustration of the phylogroup distribution throughout the infection windows which remained remarkably stable over time with no statistically significant trends (Fisher's exact test simulated *P*-value= 0.95). There were sporadic increases in the absolute numbers of various phylogroups during the peak infection periods such as phylogroup A during windows 2 and 6, as well as phylogroup B1 and phylogroup F during window 4. We collapsed the "non-peak" (*i.e.,* windows 1, 3, 5, and 7) and "peak" (*i.e.,* windows 2, 4, and 6) periods together and again found no statistically significant differences in phylogroup proportions between peak and non-peak periods (Fisher's exact test *P*-value= 0.42). The dominant B2 phylogroup accounted for 58% of infections during the non-peak windows and 47% during the peak windows (Fisher's exact test *P*-value= 0.15).

We grouped index strains by clonal complex (*i.e.,* single locus ST variants within a 7-housekeeping gene schema) to further analyze temporal trends. The leading clonal complexes (CC) were CC131 (n = 109, 44%), CC10 (n = 35, 14%), CC405 (n = 16, 7%), CC14 (n = 14, 6%), CC648 (n = 14 (6%), and CC38 (n = 10, 4%). For testing purposes, we collapsed CCs with less than 5 strains into an "uncommon" category (n = 32, 13%). Similar to our phylogroup analysis, we did not identify any significant trends for a particular clonal complex over the study

period (Fisher's exact test simulated *P*-value= 0.91). However, there were temporary absolute increases in specific CC infections within peak infection windows such as "uncommon" CCs within window 2, CC14 and CC648 within window 4, and CC10 within window 6 (**Figure 2C-D**). Importantly, CC131 remained stable in prevalence across each of the infection windows (χ^2 *P*-value = 0.9) with CC131 strains accounting for 50% of infections during non-peak windows compared to 40% of infections during peak infection windows (Fisher's exact test *P*-value= 0.15). There were no peak infection periods in which CC131 strains accounted for a higher percentage of infections relative to non-peak periods. Thus, these data indicate that peaks of ESC-R-*Ec* observed temporally over our study period were driven by a diverse combination of ESC-R-*Ec* isolates rather than expansion of CC131 strains.

Bla_{CTX-M} variants are dominant encoding genes conferring inferred ESC-R phenotype.

We searched the WGS data for genes potentially conferring the ESC-R phenotype, which we defined as those encoding ESBLs, *i.e.*, TEM, SHV, CTX-M encoding genes with a 2be Jacoby-Bush classification (12), as well as CMY and carbapenemase encoding genes. Of the 248 index isolates, 236 (95%) had at least one gene expected to confer the ESC-R phenotype with CTX-M variants being present in 220 strains (89%) and CMY variants found in 21 strains (8.4%) (*N.b.,* strains could contain multiple ESC-R encoding genes). A single CC131 strain carried an ESBL bla_{TEM-19}, two strains carried ESBL bla_{SHV-12}, one ST361 strain carried $bla_{OXA-232}$, and three CC10 strains harbored bla_{NDM-5}. The most common CTX-M encoding gene variants detected within index ESC-R-*Ec* isolates were $bla_{CTX-M-15}$ (n=134), $bla_{CTX-M-27}$ (n=29), $bla_{CTX-M-55}$ (n=26), $bla_{CTX-M-14}$ (n=19), and multiple bla_{CTX-M} variant carriage (n=7). A similar distribution was observed for recurrent ESC-R-*Ec* isolates: $bla_{CTX-M-15}$ (n=26), $bla_{CTX-M-27}$ (n=7), $bla_{CTX-M-55}$ (n=8), and $bla_{CTX-M-14}$ (n=3). Of the 12 index strains that did not carry a known ESC-R phenotype determinant, six contained bla_{OXA-1}, five encoded non-ESBL TEM enzymes, and only one isolate lacked a known exogenous β-lactamase encoding gene. The preponderance of CTX-M with a smaller role of

CMY in conferring the ESC-R phenotype in *E. coli* is consistent with previously published results in the US (9, 10).

Temporal whole genome analysis indicates low level of clonality amongst ESC-R-*Ec.*

To gain further insights into the molecular epidemiology driving the temporal dynamics of ESC-R-*Ec*, we used WGS data to analyze the population structure for all 297 isolates. The core gene inferred maximum likelihood phylogeny based on 2,695 core genes (*i.e.,* genes present in ≥99% of the 297 ESC-R *Ec*-BSI Isolates) is shown in **Figure 3**. Consistent with previous WGS studies of ESC-R *E. coli* population structure (9, 18), we observed significant variation in the degree of genetic similarity amongst strains of distinct CCs. For example, there was only an average of 28 SNPs separating a given CC14 strain from its nearest neighbor compared to 1518 for CC10 strains. Similarly, we identified several previously described subclades within the broader CC131 population (15, 20, 27). By overlying infection window with our WGS phylogeny, we sought to determine whether the increased resolution would facilitate detection of highly related strains driving infection peaks. However, in line with our phylogroup and CC-based analyses, we did not observe large numbers of phylogenetically clustered index strains occurring during infection windows (**Figure 3**). To search for clusters of closely genetically related strains more quantitatively, we determined SNP distances between all isolates from the same infection window and chose a cut-off of ≤ 30 SNPs to define closely genetically related. For the entire study period, we only identified six pairs of such strains, and only during the non-peak infection window 3 was a cluster of three strains observed. Even clonal series of isolates, such as those belonging to ST14 and subclades of ST131, were dispersed throughout the study period, occurring in both peak and non-peak windows (**Figure 3**). Consistent with the notion of similar genetic diversity being present during peak and non-peak infection windows, there was no difference in pairwise SNP distance between one isolate and its most closely related strain

during the various windows (Kruskal-Wallis test *P*-Value=0.81; Fig S2). These data indicate that even at the WGS level, peaks of ESC-R-*Ec* are driven by genetically heterogenous isolates.

ESC-R encoding genes are highly enriched within *Ec*-BSI sub-populations.

We combined the β-lactamase encoding gene data with our WGS based phylogeny to assess the relationship between *E. coli* genetic content and ESC-R mechanisms (**Figure 3**). There was a strong association of phylogroup B2 isolates harboring $bla_{\text{CTX-M-15}}$ genes as 60.4% of (81/134) $bla_{\text{CTX-M-15}}$ harboring strains were present in this phylogroup (row-wise Fisher's exact test adjusted *P*-value = 0.01). Specifically, this association was driven by the presence of $bla_{\text{CTX-M-15}}$ in 72 CC131 isolates (adj. *P*-value = 0.009), the vast majority of which are part of the previously identified $bla_{\text{CTX-M-15}}$ carrying *H*30Rx lineage (15). Similarly, there was also enrichment of the second most commonly detected CTX-M encoding gene variant, $bla_{\text{CTX-M-27}}$, in B2 index isolates (26/29 $bla_{\text{CTX-M-27}}$ harboring strains; adj. *P*-value < 0.0001) again driven by CC131 isolate carriage (23/29; Adj. *P*-value = 0.0005). Specifically, there were 19/23 $bla_{\text{CTX-M-27}}$ positive ESC-R-*Ec* belonging to the ST131-C1-M27 clade, which has a strong association with $bla_{\text{CTX-M-27}}$ carriage (27) as seen in the lower right quadrant of the core gene phylogeny (**Figure 3**). The third most prevalent CTX-M encoding gene variant, $bla_{\text{CTX-M-55}}$, was enriched within phylogroup A (12 phylogroup A strains out of 26 $bla_{\text{CTX-M-55}}$ containing index isolates; adj. *P*-value = 0.004) whereas it was only present in one B2 isolate (1/26, Adj. *P*-value < 0.0001). The final major $bla_{\text{CTX-M}}$ variant in our cohort, $bla_{\text{CTX-M-14}}$ was carried by a diverse array of genetically heterogenous strains (**Figure 3**). In contrast to the $bla_{\text{CTX-M}}$ encoding genes in which there were several large clusters of genetically adjacent strains encoding the same CTX-M variant, carriage of CMY encoding genes was sporadically observed across phylogroups. Similar to the phylogroup trend and phylogeny analyses, we did not observe any statistically significant association between ESC-R mediating enzyme variants and infection windows (χ^2 adj. *P*-value > 0.05).

Accessory genomes are distinct between B2 and non-B2 phylogroups.

We further explored genomic differences within ESC-R-*Ec* by comparing accessory genome content (*i.e.,* defined as genes carried in >5% and <95% of the total ESC-R-*Ec* population) across the full cohort. There was a total of 25,631 genes in the pan genome with 6,060 accessory genes. Through a cluster analysis using the Jaccard distance of the accessory genome between samples, we observed a strong association between phylogroup, clonal complex, and the accessory genome (**Figure 4**)**.** A neighbor joining (NJ) tree constructed from a Jaccard distance matrix shows a clear delineation of B2 and non-B2 isolates that split at the root of the tree (**Figure 4A**). Principal component analysis demonstrated clustering of accessory content by phylogroup (**Figure 4B**) and clonal complex (**Figure 4C**). Using a complimentary network graph analysis of the accessory genome (see methods) (Fig. S3), there was separation of B2 and non-B2 isolates into ten clusters highly correlating to clonal complex (**Figure 4A/4D**). Indeed, each of the clonal complexes with ≥5 isolates had at least one statistically significant association (Adj. *P*-value <0.05) with an accessory genome Markov Cluster (MCL) group (Fig. S4). Furthermore, the MCL clustering algorithm delineated the sub-clade structure of CC131 where each of the ST131-*H*30Rx subclades C1 (Cluster 2) and C2 (Cluster 3) were separated by accessory genome (**Figure 4A/4C/4D**). ST131-*H*41-O16 subclade A (Cluster 9) was found exclusively in one MCL group (**Figure 4A**) demonstrating unique accessory genome between ST131 C1/C2 and A subclades. **Figure 4E** indicates unique separation of $bla_{\text{CTX-M-27}}$ and $bla_{\text{CTX-M-55}}$ (Adj. *P*-value<0.001) within Cluster 2 (*i.e.,* CC131) and Cluster 1 (*i.e.,* CC10) respectively further highlighting differences in the clonal accessory structure of our ESC-R-*Ec* population. These analyses indicate potential differential pathways to ESC-R development dependent on phylogroup background.

Significant copy number amplifications of enriched CTX-M encoding genes occur within phylogroups

Given the growing appreciation of β-lactamase gene amplification in contributing to AMR phenotypes (28-30), we sought to characterize gene copy number variation of ESC-R genes to determine whether ESC-R gene amplification correlated with our previously identified gene enrichment associations. **Table 3** provides copy number estimates of $bla_{\text{CTX-M}}$ variants stratified by the three most common phylogroups. Tables S3 and S4 provide copy numbers of other β-lactamase encoding genes and overall estimated copy numbers of AMR genes respectively. For the 268 $bla_{\text{CTX-M}}$ genes, the median copy number variant (CNV) was 1.2× (Interquartile Range [IQR] = 1.3) with a range of 0.4 to 12.9. The median CNV across CTX-M variants was similar with $bla_{\text{CTX-M-15}}$ (n=161, CNV=1.1×; IQR =1.2), $bla_{\text{CTX-M-27}}$ (n=37, CNV=1.4×; IQR =0.8), $bla_{\text{CTX-M-55}}$ (n=37, CNV=1.5×; IQR =1.4), and $bla_{\text{CTX-M-14}}$ (n=37, CNV=1.4×; IQR =2.2) having statistically significant greater than 1 copy (one-sample Wilcoxon signed rank adj. *P*-value <0.01).

We and others have previously identified that copy number increases in narrow-spectrum and extended-spectrum β-lactamase encoding genes contribute to the development of carbapenem resistance (31, 32). Therefore, we stratified our ESC-R-*Ec* population by carbapenem susceptibility (*i.e.,* carbapenem non-susceptible vs. susceptible) to determine if there were statistically significant amplifications in carbapenem-susceptible ESC-R-*Ec* isolates. The carbapenem non-susceptible ESC-R-*Ec* had greater estimated copy numbers of $bla_{\text{CTX-M}}$ (median CNV = 2.7×; IQR = 4.2) compared to carbapenem susceptible ESC-R-*Ec* (median CNV = 1.1×; IQR = 1.0; two-sample Wilcoxon rank sum test *P*-value = 3e-7). Thus, we further characterized $bla_{\text{CTX-M}}$ gene amplification across phylogroup solely within ESC-R-*Ec* that are carbapenem susceptible (**Figure 5A**)**.** We found statistically significant amplifications for $bla_{\text{CTX-M-15}}$ and $bla_{\text{CTX-M-27}}$ in phylogroup B2 isolates and $bla_{\text{CTX-M-55}}$ in phylogroup A isolates, which

aligned with the enrichment analysis. These statistically significant trends persisted specifically within CC131 for *bla*$_{CTX-M-15}$/*bla*$_{CTX-M-27}$ and CC10 for *bla*$_{CTX-M-55}$ (Fig. S4). To assess whether *bla*$_{CTX-M}$ amplifications occurred as an adaptive response, we sought to identify isolates from the same patient that were the same clonal complex and had the same *bla*$_{CTX-M}$ variants, *i.e.,* were likely clonal, recurrent infections (**Figure 5B**). We identified 16 pairs from 14 patients in which 56% (9/16) second ESC-R *Ec*-BSI isolates were carbapenem non-susceptible following their antecedent carbapenem susceptible isolate. We found that there was a statistically significant increase (Wilcoxon Signed Rank Test *P*-value=0.03) in *bla*$_{CTX-M}$ copy numbers from the first isolate (median CNV=1.02; IQR=0.86) to the second isolate (median CNV=1.80; IQR=3.32). Notably, these *bla*$_{CTX-M}$ amplifications occurring within sequential isolates were observed for *bla*$_{CTX-M-15}$/*bla*$_{CTX-M-27}$ and *bla*$_{CTX-M-55}$ in clonal complex 131 and 10 respectively (**Figure 5B**).

F-type plasmids are primary vectors of *bla*$_{CTX-M-55}$ in non-CC131 strains.

Previous studies have focused largely on ST131 subclade associations with *bla*$_{CTX-M-15}$ and *bla*$_{CTX-M-27}$ (15, 27) including our characterization of *bla*$_{CTX-M-15}$ and *bla*$_{OXA-1}$ co-carriage on similar, multireplicon F-type plasmids contributing to carbapenem resistance (31, 32). Given the rarity of *bla*$_{CTX-M-55}$ molecular epidemiology data from North America (33), we focused on *bla*$_{CTX-M-55}$ inasmuch as it was highly prevalent throughout our sampling frame. We used a long-read approach to complete assemblies of nine *bla*$_{CTX-M-55}$ positive strains and found that most (n=7) carried *bla*$_{CTX-M-55}$ in a highly conserved IS*Ecp1* context on ~100kb IncFIB/IncFIC plasmids (**Figure 6A**). Whereas the cargo gene region was relatively variable, the F-type plasmid regions responsible for transfer, replication, and stability were highly conserved (blastn ID/coverage ≥99%). For the full cohort of *bla*$_{CTX-M-55}$ isolates (n=38), 31 isolates had IncFIB/IncFIC amplicon hits from PlasmidFinder (>80% ID;>80% COV). Short-read pileup analysis to pMB8960_1 (*i.e.,* the most conserved IncFIB/IncFIC *bla*$_{CTX-M-55}$ positive isolate), showed that 30/31 isolates had >80% breadth of coverage and >100× coverage depth for pMB8960_1.

Our long read analyses allowed for identification of *bla*$_{CTX-M-55}$ amplification arising from plasmid copy number variation (MB8960 and MB9029) as well as plasmid-to-chromosome transfer (*i.e.,* segmental duplication) from two isolates (MB6206 and MB7355) as evidenced by tracking target site duplications (TSDs) and plasmid content from IS*Ecp1-bla*$_{CTX-M-55}$ (**Figure 6B**). This included an isolate in which IS*Ecp1-bla*$_{CTX-M-55}$ disrupts the *envZ* gene, which is part of a two-component regulatory system (*i.e.,* EnvZ-OmpR) that regulates expression of outer membrane porin genes (34). A carbapenem susceptible ESC-R-*Ec* isolate (MB7355) was identified with two independent insertions within the chromosome likely originating from the IncFIB/IncFIC plasmid as evidenced through TSD tracking (**Figure 6B;** one insertion into *fhu*A gene demonstrated). Another vector of *bla*$_{CTX-M-55}$ was identified in two ST167 isolates in a substantially differing context with *bla*$_{CTX-M-55}$ upstream of IS*26*-v1 on smaller IncX1 plasmids (**Figure 6B**). The IS*26*-v1 upstream of *bla*$_{CTX-M-55}$ had truncated IS*Ecp1* (**Figure 6B;** purple striped CDS) and were within unique antimicrobial resistance gene transposons. These analyses together show that *bla*$_{CTX-M-55}$ is a major contributor to the ESC-R-*Ec* phenotype in our cohort with predominantly IncFIB/IncFIC carriage.

DISCUSSION

The past twenty years in the US have seen a consistent rise in ESC-R-*Ec* causing infection, primarily due to ESBL enzymes (2-4). While the high prevalence of CC131 observed in the US is concerning, it is not clear if these pathogens are responsible for the increased frequency of ESC-R-*Ec* (2). We sought to leverage our longitudinal collection of bloodstream isolates to address the recent temporal dynamics of *E. coli* at our tertiary care cancer center. Although we observed that CC131 strains accounted for nearly 50% of infecting isolates, non-CC131 isolates drove the temporal variability in ESC-R-*Ec* BSIs observed during the study period.

We chose to focus on BSIs because of the clear distinction between colonization and infection relative to other potential sample sites. Our large number of fully sequenced isolates and extended study timeframe provide critical information regarding contemporary molecular epidemiology of invasive ESC-R-*Ec* infections in a major US city. Although our finding of ~35% *Ec*-BSIs being ESC-R is high relative to other studies from the US (4), it is consistent with the ~40% proportion of ESC-R in *Ec*-BSI occurring in solid-organ transplant recipients (35). Whereas recent longitudinal ESC-R-*Ec* epidemiological studies from the US are lacking, we can compare our findings to those recently published from Europe (16-21). A similar theme has emerged from these investigations where for several years following introduction into the population, CC131 strains dramatically increase before becoming a stable proportion of overall *E. coli* infections (16-21). Given that CC131 had been identified as the dominant antimicrobial resistant *E. coli* CC in the US for at least 10 years prior to the start of our sampling (26, 36), our finding of a relatively stable contribution of CC131 is consistent with other studies (16-21) and suggest that, at least for our population, CC131 may have reached a stable level in the broader ESC-R-*Ec* population.

A key strength of our study was our ability to investigate the molecular epidemiology underlying the temporal variability of ESC-R-*Ec* prevalence rates. We found that ESC-R-*Ec* rates generally peaked within the last six months of the calendar year and were independent of ESC-S-*Ec* rates. Seasonality of both ESC-R-*Ec* infection and colonization have been previously reported with generally higher rates identified during warmer months (37-39). For example, a recent pooled observational cohort analysis of ESBL positive *E. coli* and *K. pneumoniae* (ESBL-E/K) carriage in the Netherlands found that the highest prevalence occurred in August (OR 1.88, 95% CI 1.02-3.49) and September (OR 2.25, 95% CI 1.30-3.89) compared to January (38). In Houston, the highest temperatures of the year are generally observed between June and September (https://www.weather.gov/hgx/climate_iah_normals_summary), which might facilitate

proliferation of *E. coli* strains leading to colonization and subsequent infection at a delayed interval. Consistent with environmental factors being a major driver of the temporal variability, we observed low genetic relatedness amongst *E. coli* strains during the peak windows of infection rather than closely related clones. On a few occasions, we identified strains from distinct patients that were closely related (*i.e.,* < 30 SNPs) consistent with hospital acquisition, but most strains were genetically diverse suggesting a community source. Given the recent success of wastewater surveillance in identifying SARS-CoV-2 and influenza transmission dynamics (40, 41), we envision that a similar approach could be used to improve understanding of how community prevalence of ESC-R pathogens impacts subsequent infections.

In previous works we have identified that non-carbapenemase producing, carbapenem resistant *Enterobacterales* (NON-CP-CRE) have increased copy numbers of genes encoding ESC-R mediating enzymes such as $bla_{\text{CTX-M}}$ variants (31, 32). To our knowledge, the data generated herein provide the first large scale, copy-number assessment of β-lactamase encoding genes mediating the ESC-R phenotype in carbapenem-susceptible *E. coli* strains. This allowed us to find that carbapenem-susceptible ESC-R-*Ec* strains have lower $bla_{\text{CTX-M}}$ copy numbers relative to carbapenem-resistant ESC-R-*Ec* strains consistent with an important role of $bla_{\text{CTX-M}}$ amplification in the transition from ESC-R to the non-CP-CRE phenotype. Moreover, via an analysis of paired isolates from the same patients, we found significant $bla_{\text{CTX-M}}$ amplifications suggesting that such amplifications are likely an adaptive mechanism to antimicrobial exposure in the host as has been observed by our group and others in the laboratory environment (31, 32, 42). Together with recently published data from ESBL-positive *E. coli* strains isolated in St. Louis (33), our long-read analyses demonstrate the diverse mechanisms by which *E. coli* can increase $bla_{\text{CTX-M}}$ encoding gene copy numbers including augmented plasmid copy numbers, multiple chromosomal copies in distinct locations, and *in situ* amplifications mediated by insertion sequences such as IS*Ecp1* and IS*26*. Although gene amplifications have long been

thought to confer a major fitness cost and thus be transient in response to antimicrobials, incorporation of *bla*$_{CTX-M}$ encoding genes into particular chromosomal locations has been found to have a low fitness cost, allowing for stable amplifications across many generations in the absence of antimicrobial pressure (43, 44). We hypothesize that the high rates of recurrence observed amongst ESC-R-*Ec* relative to ESC-S-*Ec* strains might, in part, be due to *bla*$_{CTX-M}$ amplifications occurring amongst colonizing strains in the gastrointestinal tract which would allow for their persistence during carbapenem treatment and subsequent re-infection. Such an amplification of *bla*$_{CTX-M-14}$ was recently described in an *E. coli* strain recurrently isolated from a patient's urine over many months (33).

Furthermore, we were able to identify IncFIB/IncFIC plasmids as the primary vector of IS*Ecp1*-*bla*$_{CTX-M-55}$ carriage in our cohort where amplification was occurring via both increase in plasmid copy numbers as well as plasmid-to-chromosome horizontal gene transfer. Although we did not identify a 'clonal expansion' of chromosomal IS*Ecp1*-*bla*$_{CTX-M-55}$, the independent transfer of these vectors into two different phylogroup chromosomal backgrounds suggest the potential for stable chromosomal propagation. A recent study in China found a temporal trend with increasing chromosomal *bla*$_{CTX-M-55}$ albeit without clarity to whether this was occurring within clonal populations (45). While little is known regarding *bla*$_{CTX-M-55}$ prevalence in North America (33), the enrichment of this *bla*$_{CTX-M}$ variant in the second most commonly detected phylogroup/CC in our cohort suggest these IncFIB/IncFIC::IS*Ecp1*-*bla*$_{CTX-M-55}$ plasmids are a significant contributor to ESC-R-*Ec* in our region.

Although our study has many strengths, there are important limitations. First, this was a single-center investigation at a hospital serving an immunocompromised population and thus external generalizability is unclear. This weakness is somewhat mitigated by our center serving a geographically dispersed group of patients and because the impact of ESC-R-*Ec* is particularly

severe amongst immunocompromised persons. Moreover, our finding of temporal variability amongst *E. coli* infections has been observed amongst broader populations such as a recent nationwide report from Israel (46). Second, the diverse nature of observed clonal complexes and subclones meant that there were often small numbers of non-CC131 clonal complexes during specific infection windows. Larger sampling from a more geographically dispersed area is needed to address these issues. A final weakness is that the beginning of the COVID epidemic likely led to a major disruption in infection epidemiology at our institution given the nearly 50% decrease in admissions observed in the spring and summer of 2020. Whereas both the absolute numbers of and proportion of ESC-R-*Ec* BSIs had been steadily increasing from 2017 through 2019, both significantly decreased during 2020. The reasons for these findings are not clear, but they stand in contrast to preliminary data released by the CDC indicated an increase in ESBL Enterobacterales in 2020 nationwide (47).

In summary, we present the largest whole genome-based analyses of ESC-R-*Ec* from the US analyzing recently collected strains. These data show that CC131 strains have reached a stable proportion of our ESC-R-*Ec* population. Additionally, we provide molecular insights into the temporal variability of ESC-R-*Ec* prevalence which is being increasingly appreciated but is not well understood mechanistically (16-21). The genetically diverse nature of ESC-R-*Ec* strains identified in our study indicate that efforts to mitigate serious ESC-R-*Ec* infections likely will need to address broad environmental reservoirs such as through vaccination and highlight the need to better understand ESC-R-*Ec* transmission and temporal dynamics outside of the hospital setting.

METHODS

Study design and sample collection

Escherichia coli bloodstream infections (*Ec*-BSI) occurred from March 1st, 2016, to December 31st, 2020, at MD Anderson Cancer Center (MDACC) in Houston, TX with March 1st, 2016, chosen given this was the date of implementation of the Epic Electronic Health Record system. We defined an index *Ec*-BSI isolate as first occurrence of a positive blood culture and a recurrent *Ec*-BSI isolate as a positive blood culture that occurred at least 14 days from a previous *Ec*-BSI. Antibiotic susceptibility testing of all *Ec*-BSI isolates was performed by the MDACC microbiology laboratory using the Accelerate PhenoTest™ BC (Accelerate Diagnostics), ETEST® (bioMérieux), and VITEK® 2 (bioMérieux). *Ec*-BSI Isolates were considered as extended-spectrum cephalosporin resistant (ESC-R) if they had a ceftriaxone (CRO) minimum inhibitory concentration (MIC) ≥ 4 μg/mL and/or a predicted ESBL phenotype.

Genomic sequencing

ESC-R-*Ec* isolates were sequenced by Illumina NovaSeq 6000 or Illumina MiSeq as previously described (32). There were 297 ESC-R isolates that had paired-end 150 base pair reads that passed quality control assessed using fastqc-v0.11.9. We performed long-read sequencing on an nine isolates using the Oxford Nanopore Technologies GridION platform as described previously (32).

Genome analysis and molecular typing

Quality assessed, paired-end 150 bp reads were assembled via SPAdes-v3.15.2 (48) with default parameters and the inclusion of the isolate option. Quality assessment of output SPAdes assembly contigs was performed via QUAST-v5.2.0 (49) to determine N50 and contiguity. Assemblies with (a) genomes < 4 Mb or > 6 Mb; or (b) Contigs >500 were excluded from

analysis. The EzClermont i*n silico* phylotyping tool was utilized with the 2013 Clermont PCR typing method (Waters, N EzClermont Github: https://github.com/nickp60/EzClermont). *E. coli* multilocus sequence typing (MLST) using the Achtman definition was performed with the mlst-v2.19.0 *in silico* contig scanning tool (Seemann T mlst GitHub: https://github.com/tseemann/mlst), which features the PUBMLST database (50). The COpy Number Variant quantifICation Tool (convict-v1.0) was used to identify antimicrobial resistance (AMR) genes and estimate gene copy numbers as described previously (Shropshire, W convict GitHub: https://github.com/wshropshire/convict) (32). Briefly, short-read data is used as input to KmerResistance-v2.2.0 (51, 52) to identify potential AMR genes from the resfinder database (Accessed 2022-10-01). Through a coverage depth ratio of AMR gene to housekeeping gene that accounts for variance in coverage depths, convict estimates gene copy numbers. The high-performance computing (HPC) cluster, Seadragon, that is hosted through MDACC was used to perform genomic analyses.

Pan genome analysis

Draft and complete genomes were annotated using the Prokka-v1.14.6 (53) using default parameters. Annotated gff files from prokka were used as input for pangenome analysis using Roary-v3.13.0 (54) using a protein family ID threshold of 95%. A core gene alignment was completed using Mafft-v7.508 with the Roary pipeline using a core gene threshold of 99%. This core gene alignment file was used as input to IQTree2-v2.2.0-beta to create a core gene inferred maximum likelihood phylogeny. ModelFinder was utilized to determine most appropriate DNA model with the best fit model: UNREST+FO+I+G4 chosen according to BIC. A final bootstrap consensus tree was created using 1000 SH-like aLRT replicates in addition to 1000 Ultrafast Bootstrap support values.

The accessory genome was assessed using the gene_presence_absence.csv file output from Roary. A Jaccard distance matrix was created using this binary accessory genome file after removing genes present in <5% and >95% of the total population. This distance matrix was used as input to create a neighbor joining tree using the BIONJ algorithm (55). We performed a principal component analysis of this distance matrix using base R functions. The graph network analysis was accomplished using the same gene_presence_absence.csv input using graphia (56, 57). Edge transformations using a k-nearest neighbors algorithm (n=8) were employed to reduce the number of edges from 44k to 1724. Clustering was performed using the Markov Cluster algorithm (MCL) with an inflation value=1.5 (58). Visualization of data was accomplished using ggplot2-v3.3.5 (59) and ggtree-v.3.3.1 (60).

Long read analysis

A short- and long-read assembly pipeline (Shropshire, W flye_hybrid_assembly_pipeline GitHub: https://github.com/wshropshire/flye_hybrid_assembly_pipeline) was used to close complete genomes of ONT sequenced data as described previously (32). Incomplete assemblies were re-assembled using Unicycler-v0.5.0 and manually curated for errors using short- and long-read pileups and visualizing with the integrated genome browser (IGV-v2.14.1). Comparison of plasmid vectors with BLAST ring image generator (BLAST) was performed using the Proksee server while querying AMR genes with the Comprehensive Antimicrobial Resistance Database (CARD) (61).

Statistical analysis

Prevalence rates were plotted across time using ggplot2 with a smoothing curve (*i.e.,* a loess function) to visualize prevalence trends. Time series data was assessed for monotonic trends using the Mann-Kendall Trend test and non-monotonic using the WAVK statistic. Correlations between ESC-R-*Ec* and ESC-S-*Ec* were measured with a Pearson's product moment

correlation test. We used χ^2 test (Fisher's exact when observations <5) to determine statistical differences in proportions across phylogroup/clonal complexes. A Kruskal-Wallis global test to determine differences in gene copy numbers across groups. Pairwise and one-way Wilcoxon Rank sum tests were to determine group level and single mean differences respectively with a Bonferroni correction for multiple comparisons. All statistics was computed on R (4.0.4) and α=0.05 parameter was used across all tests.

Data Availability

WGS data sequenced during this study period was submitted to NCBI and can be accessed within BioProject PRJNA924946. WGS data from previous studies can be accessed from BioProject PRJNA603908 and PRJNA836696 (32, 62). All R Scripts can be made available through request of the corresponding author.

Acknowledgements

We would like to thank the MDACC clinical microbiology lab for all their hard work in identifying, handling, and transferring these pathogenic strains of interest to us for our research projects. Core grant CA016672(ATGC) and NIH 1S10OD024977-01 grant provide funding for the Advanced Technology Genomics Core (ATGC) sequencing facility at MD Anderson Cancer Center. WCS is supported through the National Institute of Allergy and Infectious Diseases (NIAID) T32 AI141349 Training Program in Antimicrobial Resistance. Support for this study was also provided by the NIAID R21AI151536 and P01AI152999 for S.A.S. The research in the A.K. laboratory is supported by NIGMS 1R01GM133904-01 and the Welch Foundation Research Grant AU-1998-20190330. The authors acknowledge the support of the High-Performance Computing for research facility at the University of Texas MD Anderson Cancer

Center for providing computational resources that have contributed to the research results reported in this paper.

REFERENCES

1. CDC. 2019. Antibiotic Resistance Threats in the United States, 2019. U.S. Department of Health and Human Services, CDC, Atlanta, GA.
2. Jernigan JA, Hatfield KM, Wolford H, Nelson RE, Olubajo B, Reddy SC, McCarthy N, Paul P, McDonald LC, Kallen A, Fiore A, Craig M, Baggs J. 2020. Multidrug-Resistant Bacterial Infections in U.S. Hospitalized Patients, 2012–2017. New England Journal of Medicine 382:1309-1319.
3. Kaye KS, Gupta V, Mulgirigama A, Joshi AV, Scangarella-Oman NE, Yu K, Ye G, Mitrani-Gold FS. 2021. Antimicrobial Resistance Trends in Urine Escherichia coli Isolates From Adult and Adolescent Females in the United States From 2011 to 2019: Rising ESBL Strains and Impact on Patient Management. Clin Infect Dis 73:1992-1999.
4. Begier E, Rosenthal NA, Gurtman A, Kartashov A, Donald RGK, Lockhart SP. 2021. Epidemiology of Invasive Escherichia coli Infection and Antibiotic Resistance Status Among Patients Treated in US Hospitals: 2009-2016. Clin Infect Dis 73:565-574.
5. Mackinnon MC, McEwen SA, Pearl DL, Lyytikäinen O, Jacobsson G, Collignon P, Gregson DB, Valiquette L, Laupland KB. 2021. Mortality in Escherichia coli bloodstream infections: a multinational population-based cohort study. BMC Infectious Diseases 21:606.
6. Mackinnon MC, McEwen SA, Pearl DL, Parfitt EC, Pasquill K, Steele L, Laupland KB. 2021. Escherichia coli bloodstream infections in the western interior of British Columbia, Canada: a population-based cohort study. Epidemiology and Infection 149:1-28.
7. Trecarichi EM, Giuliano G, Cattaneo C, Ballanti S, Criscuolo M, Candoni A, Marchesi F, Laurino M, Dargenio M, Fanci R, Cefalo M, Delia M, Spolzino A, Maracci L, Nadali G, Busca A, Del Principe MI, Daffini R, Simonetti E, Dragonetti G, Zannier ME, Pagano L, Tumbarello M. 2019. Bloodstream infections caused by Escherichia coli in onco-haematological patients: Risk factors and mortality in an Italian prospective survey. PLOS ONE 14:e0224465.
8. Philippon A, Arlet G, Jacoby GA. 2002. Plasmid-Determined AmpC-Type β-Lactamases. Antimicrobial Agents and Chemotherapy 46:1-11.
9. Zamudio R, Boerlin P, Beyrouthy R, Madec J-Y, Schwarz S, Mulvey MR, Zhanel GG, Cormier A, Chalmers G, Bonnet R, Haenni M, Eichhorn I, Kaspar H, Garcia-Fierro R, Wood JLN, Mather AE. 2022. Dynamics of extended-spectrum cephalosporin resistance genes in Escherichia coli from Europe and North America. Nature Communications 13:7490.
10. Tamma PD, Sharara SL, Pana ZD, Amoah J, Fisher SL, Tekle T, Doi Y, Simner PJ. 2019. Molecular Epidemiology of Ceftriaxone-Nonsusceptible Enterobacterales Isolates in an Academic Medical Center in the United States. Open Forum Infectious Diseases 6:ofz353.
11. Castanheira M, Deshpande LM, Mendes RE, Canton R, Sader HS, Jones RN. 2019. Variations in the Occurrence of Resistance Phenotypes and Carbapenemase Genes Among Enterobacteriaceae Isolates in 20 Years of the SENTRY Antimicrobial Surveillance Program. Open Forum Infect Dis 6:S23-S33.
12. Bush K, Jacoby GA. 2010. Updated Functional Classification of β-Lactamases. Antimicrobial Agents and Chemotherapy 54:969-976.
13. Perez F, Endimiani A, Hujer K, Bonomo R. 2007. The continuing challenge of ESBLs. Current Opinion in Pharmacology 7:459-469.
14. Stoesser N, Sheppard AE, Pankhurst L, De Maio N, Moore CE, Sebra R, Turner P, Anson LW, Kasarskis A, Batty EM, Kos V, Wilson DJ, Phetsouvanh R, Wyllie D, Sokurenko E, Manges AR, Johnson TJ, Price LB, Peto TE, Johnson JR, Didelot X, Walker AS, Crook DW, Modernizing Medical Microbiology Informatics G. 2016.

Evolutionary History of the Global Emergence of the Escherichia coli Epidemic Clone ST131. MBio 7:e02162.

15. Price LB, Johnson JR, Aziz M, Clabots C, Johnston B, Tchesnokova V, Nordstrom L, Billig M, Chattopadhyay S, Stegger M, Andersen PS, Pearson T, Riddell K, Rogers P, Scholes D, Kahl B, Keim P, Sokurenko EV. 2013. The Epidemic of Extended-Spectrum-β-Lactamase-Producing Escherichia coli ST131 Is Driven by a Single Highly Pathogenic Subclone, H30-Rx. mBio 4:e00377-13-e00377.

16. Lipworth S, Vihta K-D, Chau K, Barker L, George S, Kavanagh J, Davies T, Vaughan A, Andersson M, Jeffery K, Oakley S, Morgan M, Hopkins S, Peto TEA, Crook DW, Walker AS, Stoesser N. 2021. Ten-year longitudinal molecular epidemiology study of Escherichia coli and Klebsiella species bloodstream infections in Oxfordshire, UK. Genome Medicine 13:1-13.

17. Royer G, Darty MM, Clermont O, Condamine B, Laouenan C, Decousser J-W, Vallenet D, Lefort A, De Lastours V, Denamur E. 2021. Phylogroup stability contrasts with high within sequence type complex dynamics of Escherichia coli bloodstream infection isolates over a 12-year period. Genome Medicine 13:77.

18. Verschuuren TD, van Hout D, Arredondo-Alonso S, Fluit AC, Reuland EA, Top J, Schurch AC, Bosch T, Bonten MJM, Kluytmans J, Willems RJL. 2021. Comparative genomics of ESBL-producing Escherichia coli (ESBL-Ec) reveals a similar distribution of the 10 most prevalent ESBL-Ec clones and ESBL genes among human community faecal and extra-intestinal infection isolates in the Netherlands (2014-17). J Antimicrob Chemother 76:901-908.

19. Kallonen T, Brodrick HJ, Harris SR, Corander J, Brown NM, Martin V, Peacock SJ, Parkhill J. 2017. Systematic longitudinal survey of invasive Escherichia coli in England demonstrates a stable population structure only transiently disturbed by the emergence of ST131. Genome Research 27:1437-1449.

20. Gladstone RA, McNally A, Pöntinen AK, Tonkin-Hill G, Lees JA, Skytén K, Cléon F, Christensen MOK, Haldorsen BC, Bye KK, Gammelsrud KW, Hjetland R, Kümmel A, Larsen HE, Lindemann PC, Löhr IH, Marvik Å, Nilsen E, Noer MT, Simonsen GS, Steinbakk M, Tofteland S, Vattøy M, Bentley SD, Croucher NJ, Parkhill J, Johnsen PJ, Samuelsen Ø, Corander J. 2021. Emergence and dissemination of antimicrobial resistance in Escherichia coli causing bloodstream infections in Norway in 2002–17: a nationwide, longitudinal, microbial population genomic study. The Lancet Microbe 2:e331-e341.

21. Rodríguez I, Figueiredo AS, Sousa M, Aracil-Gisbert S, Fernández-de-Bobadilla MD, Lanza VF, Rodríguez C, Zamora J, Loza E, Mingo P. 2021. A 21-year survey of Escherichia coli from bloodstream infections (BSI) in a tertiary hospital reveals how community-hospital dynamics of B2 phylogroup clones influence local BSI rates. Msphere 6:e00868-21.

22. Peirano G, Lynch T, Matsumara Y, Nobrega D, Finn TJ, Devinney R, Pitout JDD. 2020. Trends in Population Dynamics of Escherichia coli Sequence Type 131, Calgary, Alberta, Canada, 2006–2016. Emerging Infectious Diseases 26:2907-2915.

23. Manges AR, Geum HM, Guo A, Edens TJ, Fibke CD, Pitout JDD. 2019. Global Extraintestinal Pathogenic Escherichia coli (ExPEC) Lineages. Clin Microbiol Rev 32:e00135-18.

24. Banerjee R, Johnson JR. 2014. A new clone sweeps clean: the enigmatic emergence of Escherichia coli sequence type 131. Antimicrob Agents Chemother 58:4997-5004.

25. McNally A, Kallonen T, Connor C, Abudahab K, Aanensen DM, Horner C, Peacock SJ, Parkhill J, Croucher NJ, Corander J. 2019. Diversification of Colonization Factors in a Multidrug-Resistant Escherichia coli Lineage Evolving under Negative Frequency-Dependent Selection. mBio 10:e00644-19.

26. Johnson JR, Porter S, Thuras P, Castanheira M. 2017. The Pandemic H30 Subclone of Sequence Type 131 (ST131) as the Leading Cause of Multidrug-Resistant Escherichia coli Infections in the United States (2011-2012). Open Forum Infect Dis 4:ofx089.
27. Matsumura Y, Pitout JD, Gomi R, Matsuda T, Noguchi T, Yamamoto M, Peirano G, DeVinney R, Bradford PA, Motyl MR, Tanaka M, Nagao M, Takakura S, Ichiyama S. 2016. Global Escherichia coli Sequence Type 131 Clade with blaCTX-M-27 Gene. Emerg Infect Dis 22:1900-1907.
28. Hansen KH, Andreasen MR, Pedersen MS, Westh H, Jelsbak L, Schonning K. 2019. Resistance to piperacillin/tazobactam in Escherichia coli resulting from extensive IS26-associated gene amplification of blaTEM-1. J Antimicrob Chemother 74:3179-3183.
29. Rodriguez-Villodres A, Gil-Marques ML, Alvarez-Marin R, Bonnin RA, Pachon-Ibanez ME, Aguilar-Guisado M, Naas T, Aznar J, Pachon J, Lepe JA, Smani Y. 2019. Extended-spectrum resistance to beta-lactams/beta-lactamase inhibitors (ESRI) evolved from low-level resistant Escherichia coli. J Antimicrob Chemother 75:77-85.
30. Nicoloff H, Hjort K, Levin BR, Andersson DI. 2019. The high prevalence of antibiotic heteroresistance in pathogenic bacteria is mainly caused by gene amplification. Nat Microbiol 4:504-514.
31. Shropshire WC, Aitken SL, Pifer R, Kim J, Bhatti MM, Li X, Kalia A, Galloway-Peña J, Sahasrabhojane P, Arias CA, Greenberg DE, Hanson BM, Shelburne SA. 2021. IS26-mediated amplification of blaOXA-1 and blaCTX-M-15 with concurrent outer membrane porin disruption associated with de novo carbapenem resistance in a recurrent bacteraemia cohort. Journal of Antimicrobial Chemotherapy 76:385-395.
32. Shropshire WC, Konovalova A, McDaneld P, Gohel M, Strope B, Sahasrabhojane P, Tran CN, Greenberg D, Kim J, Zhan X, Aitken S, Bhatti M, Savidge TC, Treangen TJ, Hanson BM, Arias CA, Shelburne SA. 2022. Systematic Analysis of Mobile Genetic Elements Mediating Beta-Lactamase Gene Amplification in Noncarbapenemase-Producing Carbapenem-Resistant Enterobacterales Bloodstream Infections. Msystems 7:e00476-22.
33. Mahmud B, Wallace MA, Reske KA, Alvarado K, Muenks CE, Rasmussen DA, Burnham C-AD, Lanzas C, Dubberke ER, Dantas G. 2022. Epidemiology of Plasmid Lineages Mediating the Spread of Extended-Spectrum Beta-Lactamases among Clinical Escherichia coli. Msystems 7:e00519-22.
34. Cai SJ, Inouye M. 2002. EnvZ-OmpR Interaction and Osmoregulation in Escherichia coli. Journal of Biological Chemistry 277:24155-24161.
35. Anesi JA, Lautenbach E, Tamma PD, Thom KA, Blumberg EA, Alby K, Bilker WB, Werzen A, Tolomeo P, Omorogbe J, Pineles L, Han JH. 2021. Risk Factors for Extended-Spectrum beta-lactamase-Producing Enterobacterales Bloodstream Infection Among Solid-Organ Transplant Recipients. Clin Infect Dis 72:953-960.
36. Johnson JR, Urban C, Weissman SJ, Jorgensen JH, Lewis JS, Hansen G, Edelstein PH, Robicsek A, Cleary T, Adachi J, Paterson D, Quinn J, Hanson ND, Johnston BD, Clabots C, Kuskowski MA. 2012. Molecular Epidemiological Analysis of Escherichia coli Sequence Type ST131 (O25:H4) and blaCTX-M-15 among Extended-Spectrum-β-Lactamase-Producing E. coli from the United States, 2000 to 2009. Antimicrobial Agents and Chemotherapy 56:2364-2370.
37. Gradel KO, Nielsen SL, Pedersen C, Knudsen JD, Østergaard C, Arpi M, Jensen TG, Kolmos HJ, Søgaard M, Lassen AT, Schønheyder HC. 2016. Seasonal Variation of Escherichia coli, Staphylococcus aureus, and Streptococcus pneumoniae Bacteremia According to Acquisition and Patient Characteristics: A Population-Based Study. Infection Control & Hospital Epidemiology 37:946-953.
38. Wielders CCH, Van Duijkeren E, Van Den Bunt G, Meijs AP, Dierikx CM, Bonten MJM, Van Pelt W, Franz E, De Greeff SC. 2020. Seasonality in carriage of extended-spectrum

β-lactamase-producing Escherichia coli and Klebsiella pneumoniae in the general population: a pooled analysis of nationwide cross-sectional studies. Epidemiology and Infection 148:1-17.

39. Deeny SR, Van Kleef E, Bou-Antoun S, Hope RJ, Robotham JV. 2015. Seasonal changes in the incidence of Escherichia coli bloodstream infection: variation with region and place of onset. Clinical Microbiology and Infection 21:924-929.

40. Maryam S, Ul Haq I, Yahya G, Ul Haq M, Algammal AM, Saber S, Cavalu S. 2022. COVID-19 surveillance in wastewater: An epidemiological tool for the monitoring of SARS-CoV-2. Front Cell Infect Microbiol 12:978643.

41. Wolken M, Sun T, McCall C, Schneider R, Caton K, Hundley C, Hopkins L, Ensor K, Domakonda K, Kalvapalle P, Persse D, Williams S, Stadler LB. 2023. Wastewater surveillance of SARS-CoV-2 and influenza in preK-12 schools shows school, community, and citywide infections. Water Res 231:119648.

42. Johnson TJ, Danzeisen JL, Youmans B, Case K, Llop K, Munoz-Aguayo J, Flores-Figueroa C, Aziz M, Stoesser N, Sokurenko E, Price LB, Johnson JR. 2016. Separate F-Type Plasmids Have Shaped the Evolution of the H 30 Subclone of Escherichia coli Sequence Type 131. mSphere 1:e00121-16.

43. Yoon E-J, Gwon B, Liu C, Kim D, Won D, Park SG, Choi JR, Jeong SH. 2020. Beneficial Chromosomal Integration of the Genes for CTX-M Extended-Spectrum β-Lactamase in Klebsiella pneumoniae for Stable Propagation. Msystems 5:e00459-20.

44. Yoon E-J, Choi YJ, Kim D, Won D, Choi JR, Jeong SH. 2022. Amplification of the Chromosomal bla CTX-M-14 Gene in Escherichia coli Expanding the Spectrum of Resistance under Antimicrobial Pressure. Microbiology Spectrum:e00319-22.

45. Zeng S, Luo J, Li X, Zhuo C, Wu A, Chen X, Huang L. 2021. Molecular Epidemiology and Characteristics of CTX-M-55 Extended-Spectrum beta-Lactamase-Producing Escherichia coli From Guangzhou, China. Front Microbiol 12:730012.

46. Feldman SF, Temkin E, Wulffhart L, Nutman A, Schechner V, Shitrit P, Shvartz R, Schwaber MJ, Carmeli Y. 2022. Effect of temperature on Escherichia coli bloodstream infection in a nationwide population-based study of incidence and resistance. Antimicrobial Resistance & Infection Control 11:1-9.

47. CDC. 2022. COVID-19: US impact on antimicrobial resistance, special report 2022. U.S. Department of Health and Human Services, CDC, Atlanta, GA.

48. Prjibelski A, Antipov D, Meleshko D, Lapidus A, Korobeynikov A. 2020. Using SPAdes De Novo Assembler. Current Protocols in Bioinformatics 70:e102.

49. Mikheenko A, Prjibelski A, Saveliev V, Antipov D, Gurevich A. 2018. Versatile genome assembly evaluation with QUAST-LG. Bioinformatics 34:i142-i150.

50. Jolley KA, & Maiden, M. C. 2010. BIGSdb: scalable analysis of bacterial genome variation at the population level. BMC bioinformatics 11:595.

51. Clausen PTLC, Zankari E, Aarestrup FM, Lund O. 2016. Benchmarking of methods for identification of antimicrobial resistance genes in bacterial whole genome data. Journal of Antimicrobial Chemotherapy 71:2484-2488.

52. Clausen PTLC, Aarestrup FM, Lund O. 2018. Rapid and precise alignment of raw reads against redundant databases with KMA. BMC Bioinformatics 19.

53. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30:2068-9.

54. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MT, Fookes M, Falush D, Keane JA, Parkhill J. 2015. Roary: rapid large-scale prokaryote pan genome analysis. Bioinformatics 31:3691-3.

55. Gascuel O. 1997. BIONJ: an improved version of the NJ algorithm based on a simple model of sequence data. Molecular Biology and Evolution 14:685-695.

56. Freeman TC, Horsewell S, Patir A, Harling-Lee J, Regan T, Shih BB, Prendergast J, Hume DA, Angus T. 2022. Graphia: A platform for the graph-based visualisation and analysis of high dimensional data. PLOS Computational Biology 18:e1010310.
57. Harling-Lee JD, Gorzynski J, Yebra G, Angus T, Fitzgerald JR, Freeman TC. 2022. A graph-based approach for the visualisation and analysis of bacterial pangenomes. BMC Bioinformatics 23:1-15.
58. Van Dongen S. 2008. Graph clustering via a discrete uncoupling process. SIAM Journal on Matrix Analysis and Applications 30:121-141.
59. Wickham H. 2016. ggplot2: elegant graphics for data analysis Springer-Verlag New York
60. Yu G, Smith DK, Zhu H, Guan Y, Lam TTY. 2017. ggtree: package for visualization and annotation of phylogenetic trees with their covariates and other associated data. Methods in Ecology and Evolution 8:28-36.
61. Jia B, Raphenya AR, Alcock B, Waglechner N, Guo P, Tsang KK, Lago BA, Dave BM, Pereira S, Sharma AN, Doshi S, Courtot M, Lo R, Williams LE, Frye JG, Elsayegh T, Sardar D, Westman EL, Pawlowski AC, Johnson TA, Brinkman FS, Wright GD, McArthur AG. 2017. CARD 2017: expansion and model-centric curation of the comprehensive antibiotic resistance database. Nucleic Acids Res 45:D566-D573.
62. Shelburne SA, Kim J, Munita JM, Sahasrabhojane P, Shields RK, Press EG, Li X, Arias CA, Cantarel B, Jiang Y, Kim MS, Aitken SL, Greenberg DE. 2017. Whole-Genome Sequencing Accurately Identifies Resistance to Extended-Spectrum beta-Lactams for Major Gram-Negative Bacterial Pathogens. Clin Infect Dis 65:738-745.

TABLES

Table 1: Windows of *E. coli* Bloodstream Infection Rates per 1000 Hospital Admissions from 2016-03-01 to 2020-12-31							
Window	**1**	**2 (ESC-R Peak 1)**	**3**	**4 (ESC-R Peak 2)**	**5**	**6 (ESC-R Peak 3)**	**7**
Date Range (Days)	16/03/01 – 16/06/30 (121)	16/07/01 – 17/04/30 (394)	17/05/01 – 18/06/30 (425)	18/07/01 – 18/12/31 (183)	19/01/01 – 19/04/30 (119)	19/05/01 – 19/12/31 (244)	20/01/01 – 20/12/31 (365)
Total *E. coli* BSI per 1000 admissions	4.44	4.77	4.31	4.96	3.86	5.60	5.11
ESC-R-*Ec* per 1000 admissions	0.92	1.89	1.16	2.09	1.07	2.24	1.63
ESC-S-*Ec* per 1000 admissions	3.51	2.88	3.15	2.87	2.80	3.36	3.48

Table 2: Overview of Sequence Type Distribution Across Each Phylogroup from Sequenced Index and Recurrent ESC-R *E. coli* Isolates

Typing Group[a,b]	Index (n=248)	Recurrent (n=49)
Phylogroup B2 (n=145)	**126 (50.8%)**	**19 (38.8%)**
ST131 (n=129)	102 (80.9%)	17 (89.5%)
ST1193 (n=15)	14 (11.1%)	1 (5.3%)
Other ST (n=11)	10 (7.9%)	1 (5.3%)
Phylogroup A (n=66)	**49 (19.8%)**	**17 (34.7%)**
ST10 (n=13)	8 (16.3%)	5 (29.4%)
ST744 (n=11)	9 (18.4%)	2 (11.8%)
ST167 (n=11)	8 (16.3%)	3 (17.6%)
ST44 (n=7)	2 (4.1%)	5 (29.4%)
Other ST (n=24)	22 (44.9%)	2 (11.8%)
Phylogroup D (n=36)	**30 (12.1%)**	**6 (12.2%)**
ST405 (n=19)	16 (53.3%)	3 (50.0%)
ST38 (n=9)	7 (23.3%)	2 (33.3%)
Other ST (n=8)	7 (23.3%)	1 (16.7%)
Phylogroup F (n=20)	**19 (7.7%)**	**1 (2.0%)**
ST648 (n=15)	14 (73.7%)	1 (100.0%)
Other ST (n=5)	5 (26.3%)	0 (0.0%)
Phylogroup B1 (n=20)	**16 (6.5%)**	**4 (8.2%)**
ST224 (n=6)	5 (31.3%)	1(25.0%)
Other ST (n=13)	11 (57.9%)	3 (75.0%)
Phylogroup C (n=9)	**7 (2.8%)**	**2 (4.1%)**
ST410 (n=7)	5 (71.4%)	1 (50.0%)

Other ST (n=2)	2 (28.6%)	1 (50.0%)

[a]'Other ST' consists of sequence types detected at ≤5 total counts within each phylogroup.

[b]Phylogroup G consists of one observation (ST501) thus was not included in table.

757

Table 3: Copy Number Variants of β-lactamase Encoding Genes

	CTX-M-15	CTX-M-27	CTX-M-55	CTX-M-14	Other CTX-M
Total					
Count	161	37	37	23	10
Mean (SD)	1.62 (1.29)	1.87 (1.89)	2.50 (2.65)	2.61 (2.57)	3.55 (3.35)
Median (IQR)	1.11 (1.19)	1.40 (0.80)	1.50 (1.43)	1.42 (2.21)	1.31 (5.36)
Range	0.37-8.72	0.42-9.10	0.62-12.9	0.64-10.8	0.60-9.45
Phylogroup B2					
Count	92	33	1	6	3
Mean (SD)	1.41 (0.96)	1.12 (0.36)	3.41	2.53 (2.06)	4.80 (3.66)
Median (IQR)	1.03 (0.54)	1.20 (0.354)	NA	1.75 (1.83)	6.54 (3.33)
Range	0.44-5.84	0.42-9.10	NA	0.88-6.33	0.60-7.26
Phylogroup A					
Count	31	3	19	5	2
Mean (SD)	1.66 (1.72)	1.12 (0.36)	2.39 (2.09)	1.65 (1.42)	1.08 (0.13)
Median (IQR)	1.07 (0.97)	1.20 (0.35)	1.68 (1.82)	1.01 (0.31)	1.08 (0.09)
Range	0.50 - 8.72	0.72 - 1.44	0.62 - 7.20	0.78-4.16	0.98-1.17
Phylogroup D					
Count	21	1	5	4	3
Mean (SD)	2.02 (1.48)	1.8	1.25 (0.49)	1.80 (0.98)	3.90 (4.81)
Median (IQR)	2.00 (1.62)	NA	1.30 (0.53)	1.43 (0.54)	1.31 (4.25)
Range	0.37-5.14	NA	1.00-1.96	1.10-3.26	0.94-9.45

SD = Standard Deviation; IQR = Interquartile Range; Copy Number Variants are estimated based on CONVICT results

758

FIGURES

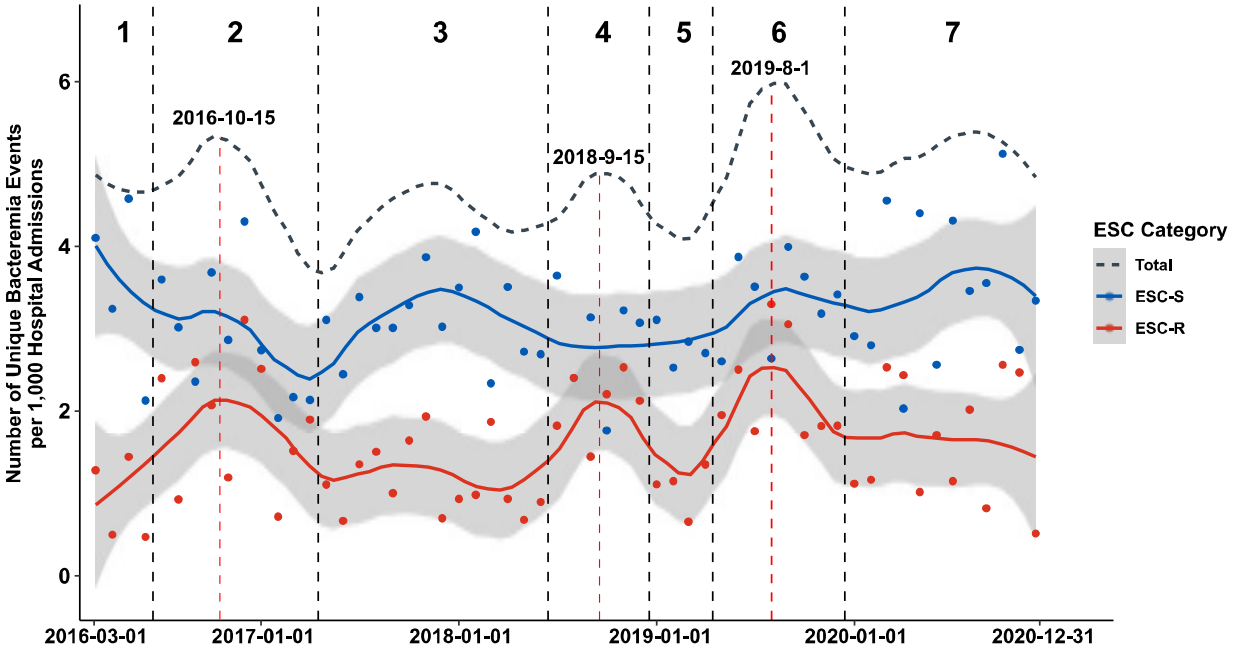

Figure 1. Number of Bacteremia Events Observed per Month from March 2016 to December 2020 at MD Anderson Cancer Center. Number of index *E. coli* bacteremia events per 1000 hospital admissions for both ESC-R (red) and ESC-S (blue) isolates. A local regression curve for each respective ESC category is provided with 95% confidence intervals (gray area). The dotted gray line represents the prevalence trend for total *E. coli* bacteremia events. Windows of peak infection estimates were observed in windows 2, 4, and 6 as demarcated above by vertical dotted black lines with estimated peak date indicated within peak regions (vertical dotted red lines).

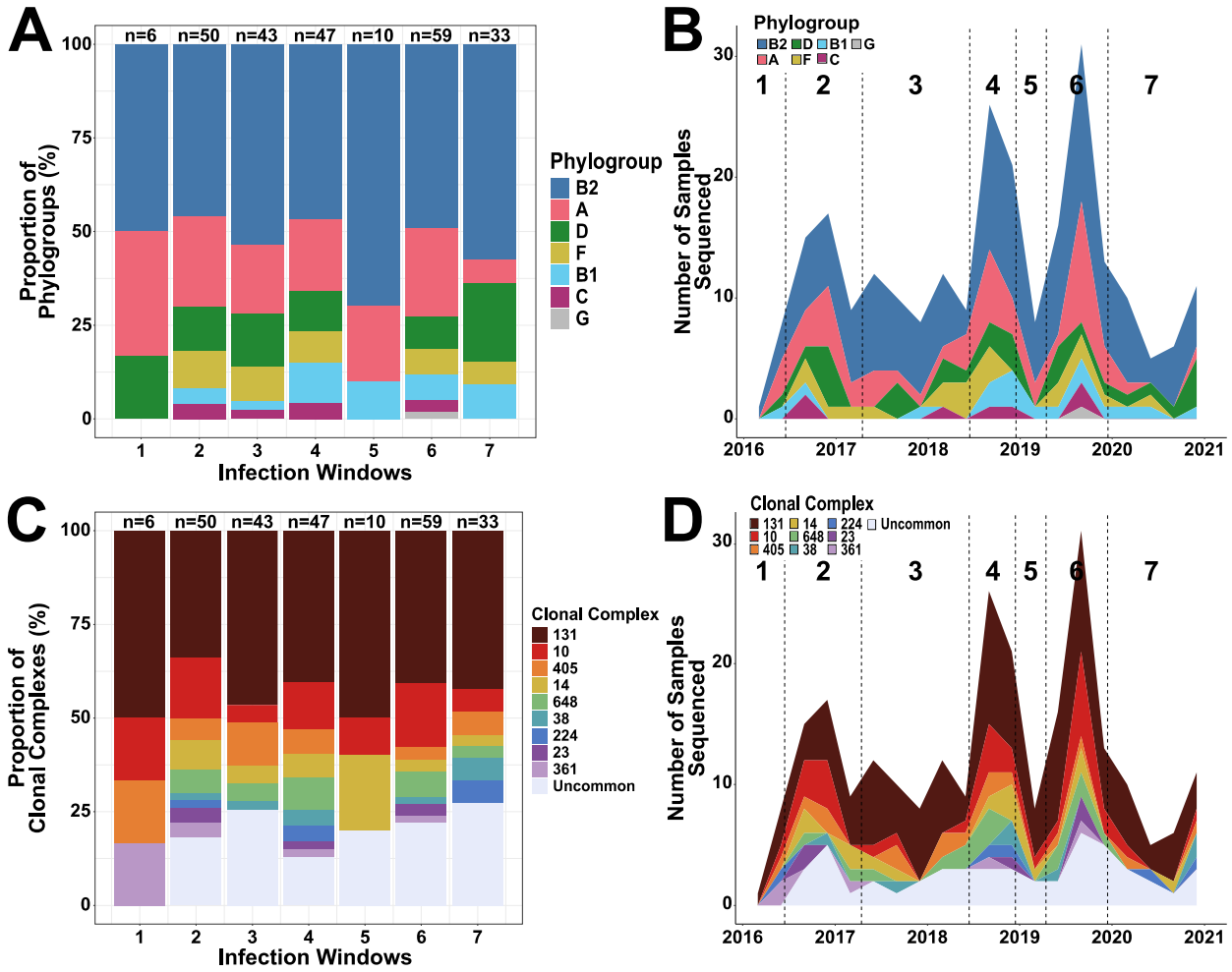

Figure 2. Temporal Distribution of ESC-R *Ec*-BSI Isolates. The proportional distribution and absolute frequency counts by phylogroup **(A-B)** and clonal complex **(C-D)** respectively across each infection window. **(A-C)** Group totals are listed above each of the windows. **(B-D)** Data shown are line graphs connecting month-to-month frequency counts colored by proportions of phylogroup **(B)** and clonal complex **(D)** as indicated in the legend.

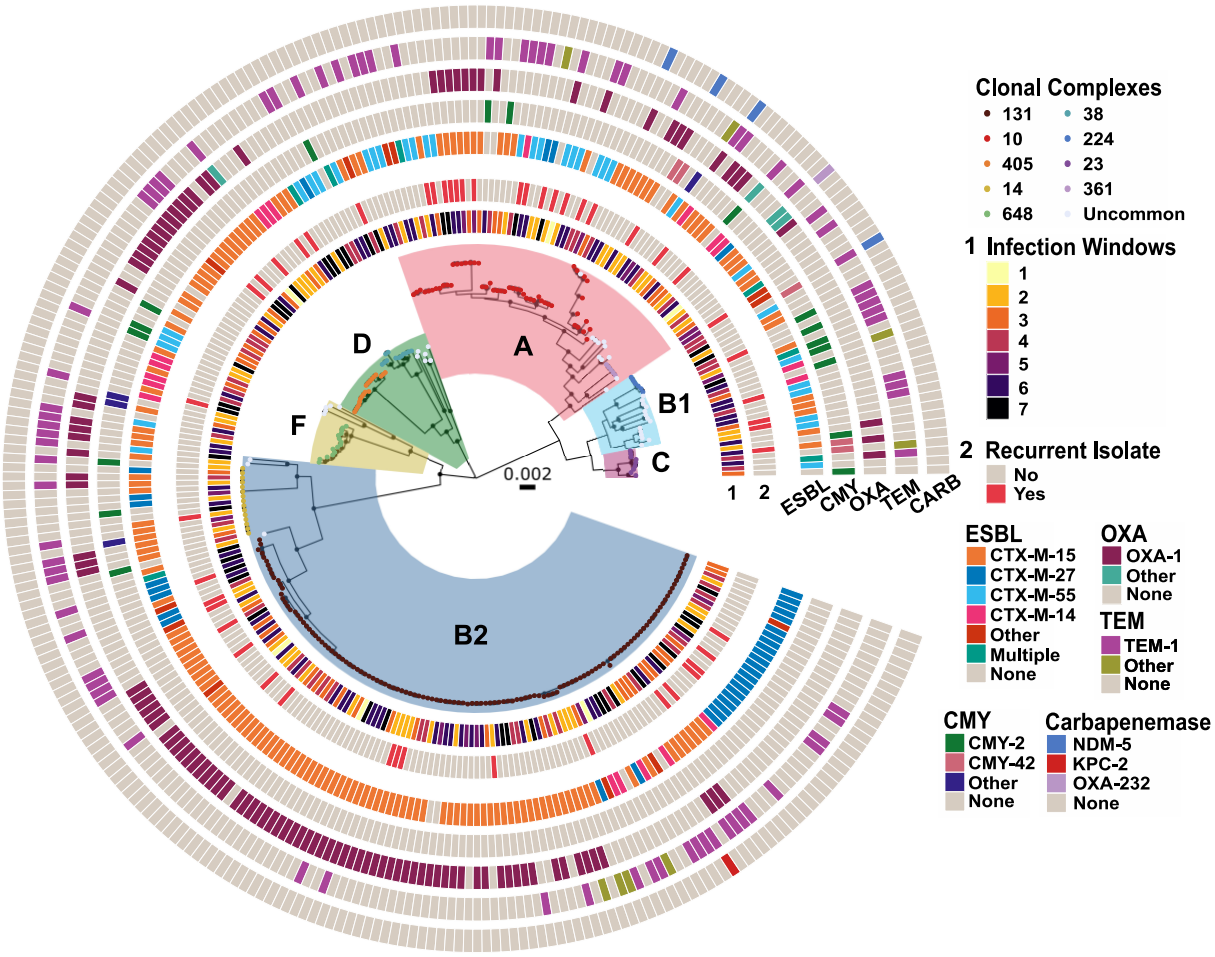

Figure 3. Population Structure of ESC-R *Ec*-BSI isolates. A core gene alignment inferred maximum likelihood phylogenetic tree with phylogroups highlighted within each monophyletic group. Branch tip colors indicate clonal complex number as shown in the figure inset with clonal complexes observed <5 counts grouped as “Uncommon”. Rings corresponding to infection window (1), recurrence (2), and β-lactamases are labelled respectively. Internal nodes labelled with black circles have SH-aLRT ≥ 80% and UFboot ≥ 95%.

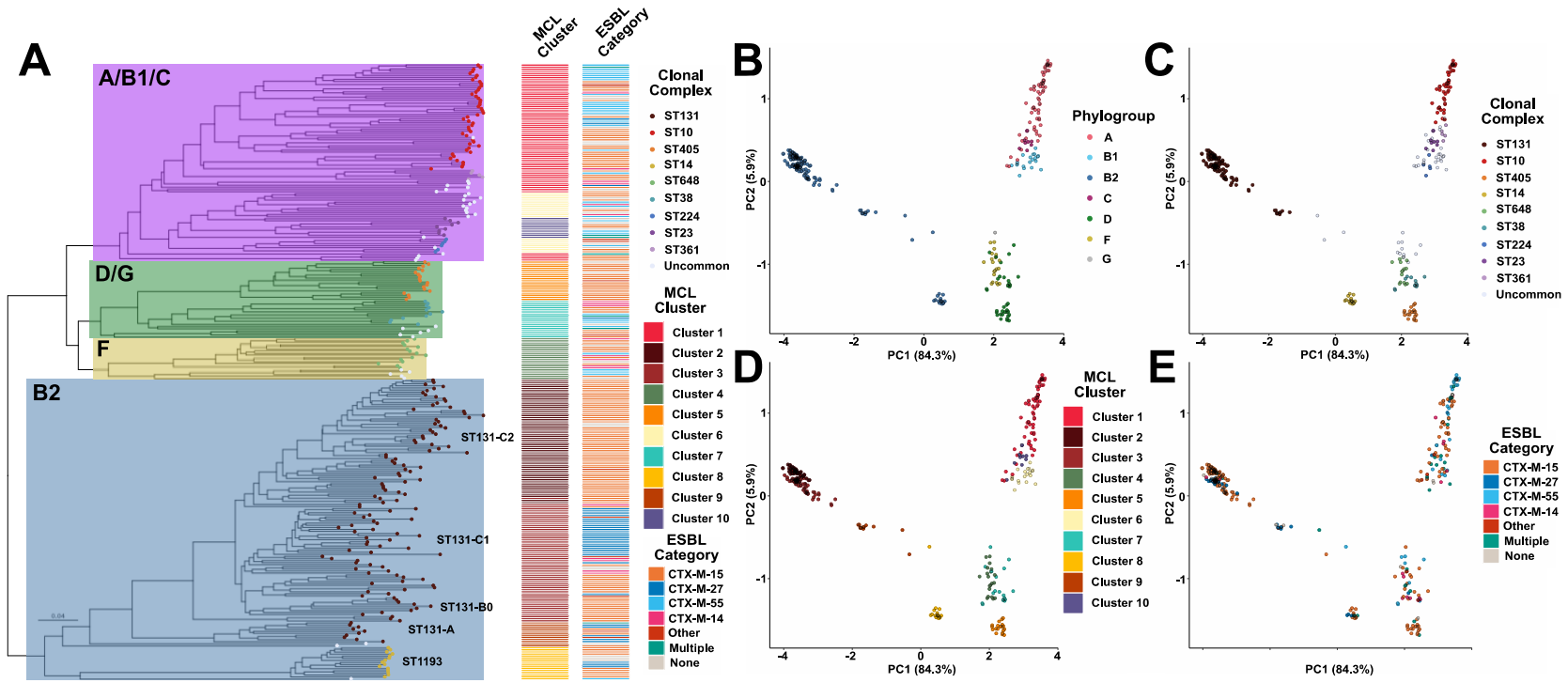

Figure 4. Accessory Genome Clustering by ST131 and non-ST131 ESC-R Isolates. (A) Jaccard distance matrix inferred NJ tree of accessory genome content with metadata indicated in legend. Phylogroups and chosen STs/clades are highlighted within tree and labelled accordingly. Principal components, PC1 and PC2, of accessory genome content are plotted with percent variance explained and stratified by **(B)** phylogroup, **(C)** clonal complex, **(D)** MCL cluster, **(E)** ESBL status indicated in respective legends.

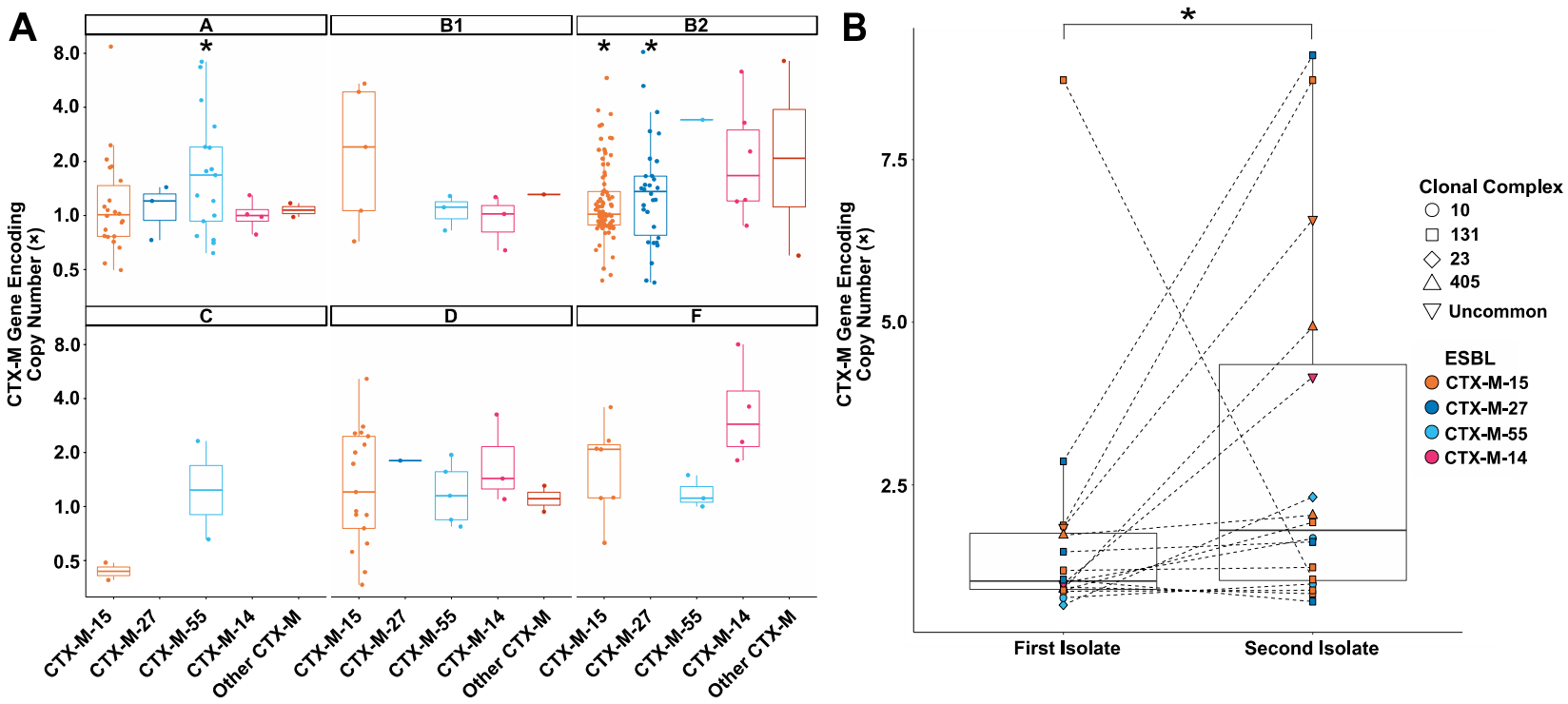

Figure 5. *Bla*$_{\text{CTX-M}}$ Amplification Occurring within ESC-R-*Ec* Isolates. (A) Copy number estimates of *bla*$_{\text{CTX-M}}$ variants stratified by phylogroup which is indicated in box. One-sample Wilcoxon rank sum test for amplification (*i.e.*, copy number > 1) performed across each *bla*$_{\text{CTX-M}}$ variant with * indicating *P*-value<0.05 **(B)** Estimated copy number of *bla*$_{\text{CTX-M}}$ variants of paired ESC-R serial isolates (selection criteria as outlined in the text). Shapes and colors indicate clonal complex and CTX-M variant as noted in the legend. Dotted lines connect paired isolates. Paired Wilcoxon-signed rank test was statistically significant with * indicating *P*-value<0.05.

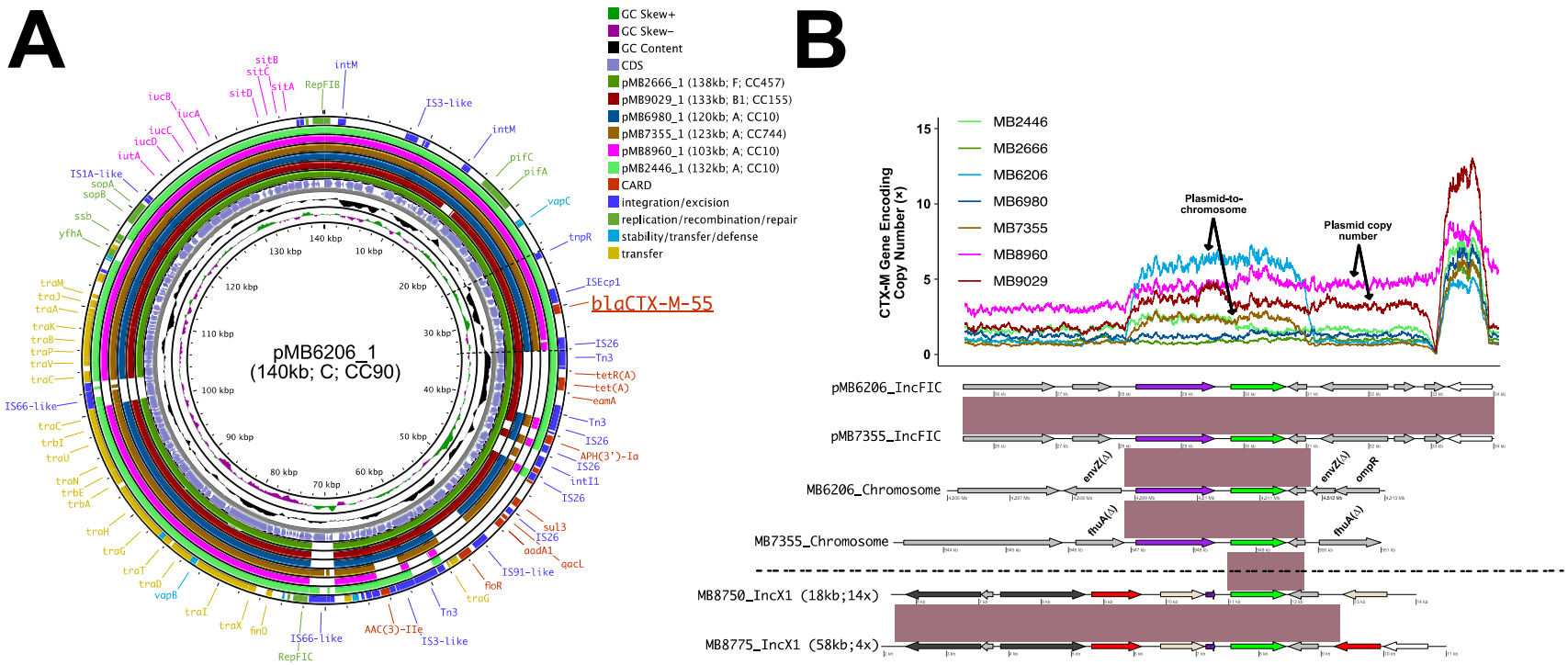

Figure 6. Characterization of mobile genetic elements associated with transmission and amplification of *bla*$_{\text{CTX-M-55}}$ within non-B2 isolates. (A) BRIGs output of similar IncFIB/IncFIC plasmids with conserved carriage of IS*Ecp1-bla*$_{\text{CTX-M-55}}$ carriage (Underlined) in 4 different non-B2 phylogroups. Inner ring indicates GC content/skew with coding DNA sequences in reverse/forward orientation. Next seven rings indicate blastn identities to pMB6206_1 as indicated in legend. Outer ring includes plasmid, mobile genetic element, and antimicrobial resistance gene content as identified by the Comprehensive Antimicrobial Resistance Database (CARD). **(B)** Blast nucleotide and gene amplification comparisons of differential *bla*$_{\text{CTX-M-55}}$ vectors responsible for plasmid

transmission as well as potential vertical transmission via chromosomal insertion. Normalized coverage depth of conserved ~10kb region that harbors IS*Ecp1*-*bla*$_{CTX-M-55}$ (~10kb) as demarcated by dotted black lines in **(A).** Inferred plasmid copy number vs. plasmid-to-chromosome events as indicated by tracking target site duplications of IS*Ecp1*-*bla*$_{CTX-M-55}$ on complete assemblies as well as mean coverage depths of full-length plasmid. Coding sequences within blastn comparisons are identified as followed: IS*Ecp1* (purple), *bla*$_{CTX-M-55}$ (green), AMR genes (red), IS*26* (white), IS*26*-v1 (beige), other MGE (black), other CDS (gray). Black horizontal line demarcates IncFIB/IncFIC positive isolates from high copy number IncX1 plasmids harboring *bla*$_{CTX-M-55}$ from ST167 isolates. Parentheticals following plasmid label indicate plasmid length and copy number. CDS with striped, purple label indicates IS*Ecp1* truncated by IS*26*-v1 upstream of *bla*$_{CTX-M-55}$.