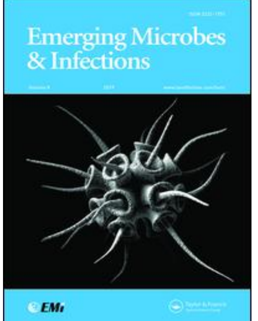


Emerging Microbes & Infections

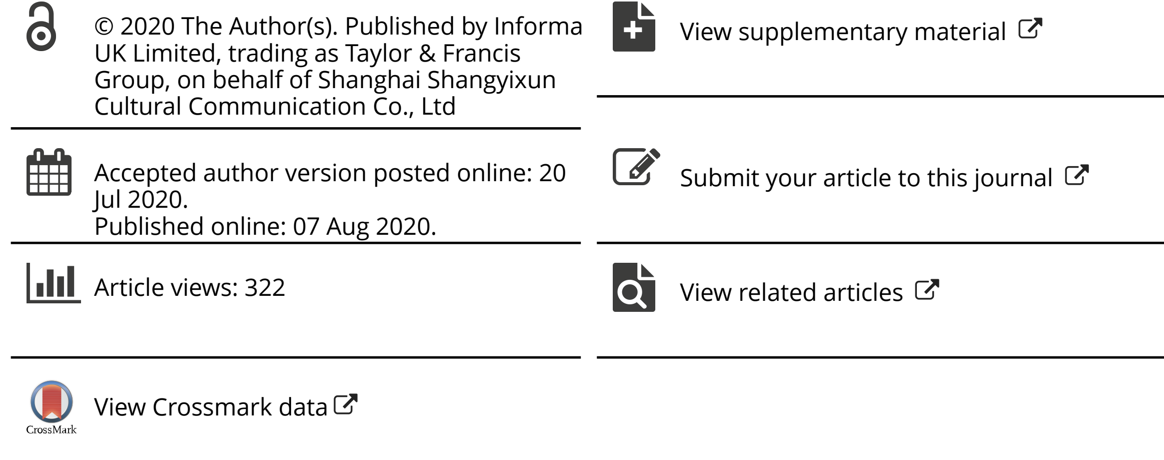
ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/temi20

Escherichia coli ST8196 is a novel, locally evolved, and extensively drug resistant pathogenic lineage within the ST131 clonal complex

Priyanka Hastak , Mathieu Fourment , Aaron E. Darling , Thomas Gottlieb , Elaine Cheong , John Merlino , Garry S. A. Myers , Piklu Roy Chowdhury & Steven P. Djordjevic

To cite this article: Priyanka Hastak , Mathieu Fourment , Aaron E. Darling , Thomas Gottlieb , Elaine Cheong , John Merlino , Garry S. A. Myers , Piklu Roy Chowdhury & Steven P. Djordjevic (2020) *Escherichia coli* ST8196 is a novel, locally evolved, and extensively drug resistant pathogenic lineage within the ST131 clonal complex, Emerging Microbes & Infections, 9:1, 1780-1792, DOI: 10.1080/22221751.2020.1797541

To link to this article: https://doi.org/10.1080/22221751.2020.1797541



© 2020 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group, on behalf of Shanghai Shangyixun Cultural Communication Co., Ltd

Accepted author version posted online: 20 Jul 2020.
Published online: 07 Aug 2020.

View supplementary material

Submit your article to this journal

Article views: 322

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=temi20

Emerging Microbes & Infections
2020, VOL. 9
https://doi.org/10.1080/22221751.2020.1797541

Escherichia coli ST8196 is a novel, locally evolved, and extensively drug resistant pathogenic lineage within the ST131 clonal complex

Priyanka Hastak[a], Mathieu Fourment[a], Aaron E. Darling[a], Thomas Gottlieb[c,d], Elaine Cheong[c,d], John Merlino[c,d], Garry S. A. Myers[a], Piklu Roy Chowdhury[a,b] and Steven P. Djordjevic[a,b]

[a]The ithree institute, University of Technology Sydney, Ultimo, Australia; [b]Australian Centre for Genomic Epidemiological Microbiology, University of Technology Sydney, Broadway, Australia; [c]Department of Microbiology and Infectious Diseases, Concord Hospital, Concord, Australia; [d]Faculty of Medicine, University of Sydney, Sydney, Australia

ABSTRACT

The *H*30Rx subclade of *Escherichia coli* ST131 is a clinically important, globally dispersed pathogenic lineage that typically displays resistance to fluoroquinolones and extended spectrum β-lactams. Isolates EC233 and EC234, variants of ST131-*H*30Rx with a novel sequence type (ST) 8196, isolated from unrelated patients presenting with bacteraemia at a Sydney Hospital in 2014 are characterised here. EC233 and EC234 are phylogroup B2, serotype O25:*H*4A, and resistant to ampicillin, amoxicillin, cefoxitin, ceftazidime, ceftriaxone, ciprofloxacin, norfloxacin and gentamicin and are likely clonal. Both harbour an IncFII_2 plasmid (pSPRC_Ec234-FII) that carries most of the resistance genes on an IS*26* associated translocatable unit, two small plasmids and a novel IncI1 plasmid (pSPRC_Ec234-I). SNP-based phylogenetic analysis of the core genome of representatives within the ST131 clonal complex places both isolates in a subclade with three clinical Australian ST131-*H*30Rx clade-C isolates. A MrBayes phylogeny analysis of EC233 and EC234 indicates ST8196 share a most recent common ancestor with ST131-*H*30Rx strain EC70 isolated from the same hospital in 2013. Our study identified genomic hallmarks that define the ST131-*H*30Rx subclade in the ST8196 isolates and highlights a need for unbiased genomic surveillance approaches to identify novel high-risk MDR *E. coli* pathogens that impact healthcare facilities.

ARTICLE HISTORY Received 12 April 2020; Revised 13 July 2020; Accepted 14 July 2020

KEYWORDS *Escherichia coli* ST8196; *E. coli* ST131 clonal complex; ST131- *H*30Rx; $bla_{CTX-M-15}$; *qnrB4*

Introduction

Resistance to last resort antibiotics severely restricts treatment options for infections caused by extra-intestinal pathogenic *Escherichia coli* (ExPEC) lineages such as sequence type (ST) 131 [1,2]. Although *E. coli* ST131 is most frequently reported from human infections, isolates have been recovered from food-producing animals, mammals, birds, cats, dogs, sheep, agriculture/horticulture, water and soil [3–7]. ST131 is a globally dominant, pandemic and multi-drug resistant (MDR) lineage, associated with both community-acquired (CA) and hospital-acquired (HA) bloodstream infections (BSIs) [8–12]. It comprises three major ancestral clades (A, B and C), with clade C representatives predominating clinical ST131 isolates globally [13–18]. Interclade diversity is primarily attributed to recombination events that have shaped representative genomes [15]. Isolates clustering in these three clades are characterized by the presence of specific type I fimbrial adhesin gene (*fimH*) alleles, and hence are widely recognized as ST131 -A/*H*41, -B/*H*22 and -C/*H30*. Clade C is predicted to have originated from a single progenitor prior to 2000 and diverged into three distinct sub-clades -C0, C1 and C2. ST131 sub-clades C1 and C2 carry diverse IncF plasmids and often host genes that render critically important antibiotics ineffective [2,13,16,19]. Strains in subclade C2 are frequently extensively drug-resistant (defined as *H*30Rx), and characterized by the presence of IncF2:A1:B plasmid replicons with a $bla_{CTX-M-15}$ gene encoding extended-spectrum β-lactamase. Subclade C1 isolates carry IncF1:A2:B20 plasmids and are usually not associated with $bla_{CTX-M-15}$ [14,19]. Insertion element IS*Ecp1* mobilizes $bla_{CTX-M-15}$ and is commonly identified adjacent to it on IncF2:A1:B plasmid replicons in the *H*30Rx subclade. Occasionally the IS*Ecp1*-$bla_{CTX-M-15}$ module is also located on chromosomes of highly resistant *H*30Rx/subclade C2 strains [15,16]. Fluoroquinolone resistance is another defining feature of ST131 clades C1 and C2 and is attributed to

CONTACT Piklu Roy Chowdhury piklu.bhattacharya@uts.edu.au The ithree institute, University of Technology Sydney, City Campus, Ultimo, NSW 2007, Australia Australian Centre for Genomic Epidemiological Microbiology, University of Technology Sydney, PO Box 123, Broadway, NSW 2007, Australia; Steven P Djordjevic Steven.Djordjevic@uts.edu.au The ithree institute, University of Technology Sydney, Ultimo, Australia; Australian Centre for Genomic Epidemiological Microbiology, University of Technology Sydney, Broadway, Australia

Supplemental data for this article can be accessed at https://doi.org/10.1080/22221751.2020.1797541

© 2020 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group, on behalf of Shanghai Shangyixun Cultural Communication Co., Ltd
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

chromosomally located alleles *gyrA1*AB and *parC1a*AB [14].

Core genomes of ST131 are highly conserved and typically comprise ≥3150 genes. Variation in resistance and virulence profiles are driven by differences in the accessory genome [20,21]. A recent pangenome analysis of a global collection of 4071 ST131 isolates predicted the potential for outbreaks caused by close genetic variants with altered resistance gene cargoes [14], highlighting the need for meticulous genomic surveillance of this globally dispersed lineage. Close genetic variants of isolates belonging to a specific multi-locus sequence type (MLST, abbreviated as ST) are usually grouped as clonal complexes (CC) and comprise genetically similar isolates exhibiting a single change in the allelic profile of genes that define the central MLST genotype [20]. At the time of investigation, the ST131CC comprised 99 different STs in the Enterobase database (https://enterobase.warwick.ac.uk/species/index/ecoli), primarily sourced from human infections. Emergence of genetic variants of ST131 forming ST131-clonal groups have been reported in the past [22]; however, to the best of our knowledge, there are no studies that seek to characterize the differences and evolutionary relationships among STs residing within the ST131CC.

Here, we characterized two isolates, EC233 and EC234, that belong to a novel sequence type – ST8196, within the ST131CC. The two ST8196 isolates were collected from unrelated patients treated for bacteraemia at Concord Repatriation Hospital in Sydney in November 2014 [23]. They are both resistant to third-generation amino-penicillins (ampicillin and amoxicillin), second and third-generation cephalosporins (cefoxitin, ceftazidime and ceftriaxone), quinolones (ciprofloxacin and norfloxacin), and the aminoglycoside antibiotic gentamicin, which establishes their "extensively drug resistant" [24] status. We examined the genetic relatedness between these ST8196 isolates with their closest relatives in the ST131CC and assessed their evolutionary origin. Our comprehensive characterization of the two genomes suggest that ST8196 has likely evolved from representatives of MDR ST131-*H*30Rx clones circulating in Australia.

Material and methods

Bacterial strains and growth conditions

E. coli EC233 and EC234 were sub-cultured on Lysogeny Broth (LB) agar plates or broth and grown at 37°C for 18 hours. Genomic DNA was extracted from overnight broth cultures grown with aeration on a rotary shaker set at 2500 rpm. Antibiotic susceptibility profiles were generated using the VITEK2 (BioMérieux) system and interpreted using the European Committee on Antimicrobial Susceptibility Testing (EUCAST) breakpoints [25]. Synergy tests were used to phenotypically confirm production of extended spectrum beta lactamase (ESBL) enzymes.

Genomic DNA extraction, library preparation and whole-genome sequencing:

For Illumina HiSeq sequencing, genomic DNA was extracted using 2 ml of overnight culture with the ISOLATE II Genomic DNA kit (Bioline, Australia), following the manufacturer's protocol, and quantified using the Qubit fluorometer and dsDNA HS Assay kit (Thermo Fisher Scientific, Australia). Whole-genome sequencing libraries were prepared from 2 ng of gDNA using a previously described variation of the Illumina Nextera protocol [26], and sequenced on Illumina HiSeq 2500 v4 sequencer (Illumina, San Diego, CA, USA) in PE150 mode.

Isolate EC234 was also sequenced on the PacBio Single Molecule Real-Time (SMRT) sequencing platform. Genomic DNA for SMRT sequencing was isolated from 1.8 ml of overnight cultures using a MoBio UltraClean Microbial DNA isolation kit (Qiagen, Germany), following the manufacturer's protocol. Sequencing was performed at the Ramaciotti Center for Genomics, University of New South Wales (Sydney, Australia). Demultiplexed PacBio long reads were assembled with the Illumina short reads, using the Unicycler hybrid genome assembler [27] and submitted in GenBank under accession number WMLE00000000. The Illumina-only assembly of EC233 is available in GenBank under accession number SAMN10187599.

Genotypic profiling and genome-wide comparison of isolates

The new sequence type designation for EC233 and EC234 were acquired from the Enterobase database (http://enterobase.warwick.ac.uk/species/index/ecoli). The resfinder, virulence finder, PlasmidFinder, serotype finder and CHtyping databases were accessed via the Centre of Genomic Epidemiology website (http://www.genomicepidemiology.org/). Preliminary genome annotations were generated using an online version of RASTtk [28]. Putative antimicrobial resistance genes of interest were confirmed using stand-alone BLASTn analyses and an in-house database on the high-power computing cluster available at the University of Technology Sydney, prior to manual verification of genes of interest using NCBI-ORF finder. Iterative BLASTn and BLASTp searches [29–31]were used for confirmation of plasmid annotations. Only ORFs showing >95% sequence similarity across 100% of the query sequence were considered for further analysis. Figures comparing BLASTn alignment of genome and plasmid sequences were generated using BRIG v0.95 [32] set at

90% lower and 100% upper identity thresholds, SnapGene v3.3.4 (GSL Biotech) and EasyFig version 2.2.2 [33]. Efflux pumps associated with drug resistance in *E. coli* were identified using the Resistance Gene Identifier (RGI) module in the online version of the Comprehensive Antibiotic Resistance Database (CARD) (https://card.mcmaster.ca/). Genomic islands of EC234 were identified by island-viewer software (https://www.pathogenomics.sfu.ca/islandviewer/).

Roary [34]was used for Pan Genome analysis. The binary tree for the presence and absence of accessory genes were recalculated using the neighbour-joining algorithm and a Jaccard dissimilarity matrix from the binary data file containing the presence and absence of gene clusters produced by Roary. Support values were derived using non-parametric bootstrapping with 1000 replicates. This approach was implemented in a *R* script [35]. All figures were edited using Adobe PhotoShop CS6 (Version 13.1.2).

Phylogenetic analysis of ST8196 isolates

The SNP phylogenies of genomes and plasmids were inferred with parSNP, using the -c and -x flags to evoke forced alignment across collinear blocks and recombination filtering to improve accuracy of the estimated clonal ancestry [36]. The completely closed and annotated *E. coli* EC958 ST131 (NZ_HG941718.1) genome [37] was used as a reference in all SNP based analyses. For the ST131CC phylogeny, 72% of the reference genome EC958 was found conserved among the cohort of 87 genomes and aligned to calculate phylogenetic inferences. Twenty-four thousand and eleven variant sites were utilized in the construction of the phylogenetic tree presented in Figure 1. For the ST131 and ST8196 phylogeny, initially 621 ST131 genomes from geographically distinct areas and an isolation date were downloaded from Enterobase and aligned using parSNP. Fifty-four percent of the EC958 reference genome was conserved among all 621 aligned genomes, containing 4614 variant sites. A maximum-likelihood tree was constructed using RAxML [38] and the temporal signal was assessed using TempEst [35]. The evolutionary history of the genomes was hypothesized using a Bayesian Markov chain Monte Carlo method as implemented in the MrBayes package (version 3.2.7a) [39] on a subset of 148 genomes aligned over 2,780,909 nt and used for phylogenetic inferences presented in this manuscript. The GTR substitution model with the gamma model of rate heterogeneity (4 categories) was utilized. An unconstrained compound gamma-Dirichlet prior was placed on the branch lengths for the tree and two independent chains were run for 30 million iterations each. The convergence was checked using the average standard deviation of split frequencies (ASDSF) diagnostics [39]. Phylogenetic tree figures were generated using FigTree (http://tree.bio.ed.ac.uk/software/figtree/) and compiled with metadata using online version of iToL (https://itol.embl.de/) [40].

Results

EC233 and EC234 are representatives of ST8196, a novel sequence type within ST131-H30Rx clonal complex

In silico multi-locus sequence typing (MLST) of EC233 and EC234 led to the establishment of a novel sequence type, ST8196, within the ST131 clonal cluster. Both genomes typed as phylogroup B2 and serotype O25:H4. Of the seven genes used for Achtman *E. coli* MLST profiling, ST131 is characterized by allelic variant number 13 of the *icd* gene encoding iso-citrate dehydrogenase while ST8196 had allelic variant 912. Although 99 STs formed the ST131CC in Enterobase, genome sequences representing 70 STs were available for downloading at the time of investigation. Our SNP based phylogenetic tree, with representatives of 70 STs within ST131CC, resolved into three major clades (I, II and III), akin to ST131 population structure, with reference strains of ST131 sub-clades A, B and C [15,41] clustering within clade I, II and III respectively (Figure 1). Clade III comprised 44 ST representatives including the ST8196 representatives, ST131 clade C reference genome S115EC_cladeC-ESC_CA5436 and EC958. The genotypic profile of S115EC_cladeC-ESC_CA5436 described previously [15] confirmed the presence of the *fimH30* allele, chromosomal fluoroquinolone-resistant genes *gyrA1*AB and *parC1a*AB and $bla_{\text{CTX-M-15}}$ and a representative of the *H*30Rx lineage. Isolate S115EC_cladeC-ESC_CA5436 was collected from Brisbane in 2011 and EC958 [37] is the closed ST131 reference genome [15] available in GenBank. Our observations were based on parSNP alignments, where 72% of the core genome of EC958 was aligned to infer phylogenetic relationships and 24,011 variant sites were used in the construction of the tree.

The ST8196 isolates clustered into a smaller subclade within clade III comprising five genomes (red box and inset in Figure 1), including Sydney isolate EC70, S115EC_cladeC-ESC_CA5436 and a ST8189 isolate ESC_RA8813AA. While the ST8196 isolates EC233 and EC234 in the subclade grouped with EC70, that was collected in 2013 from Sydney, the ST8189 isolate grouped with the Brisbane isolate S115EC_cladeC-ESC_CA5436. Six hundred and seventy-two variant sites were considered in resolving the ancestral relationships of isolates in this subclade of five genomes. Using pairwise comparison of regions of genomes aligned in the parSNP phylogeny analysis,

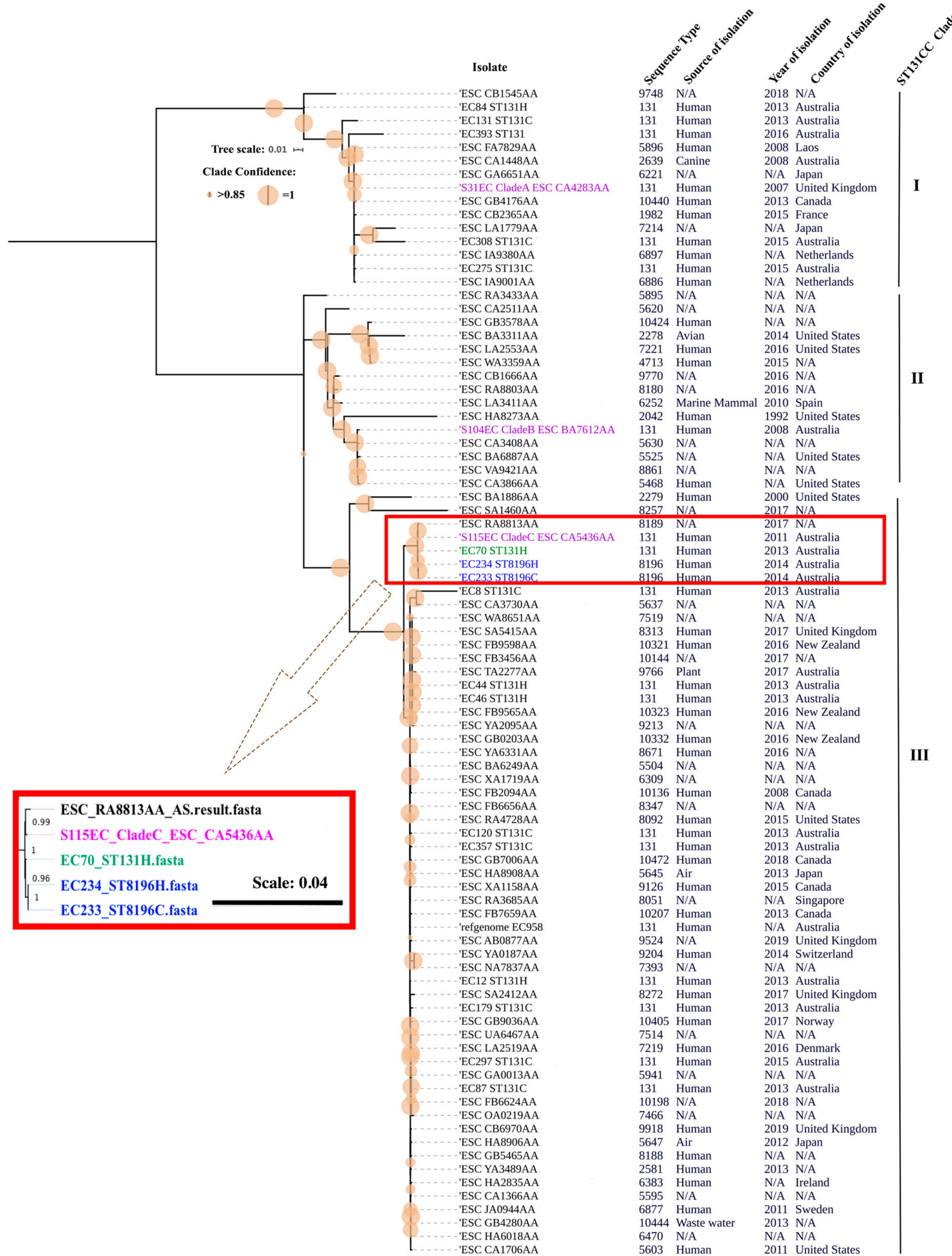


Figure 1. SNP phylogeny of the core genomes of sequence type representatives comprising the ST131 clonal cluster (ST131CC). Tree scale represents substitution per site. The tree also includes ST131 isolates from our extended collection [33], ST131 reference genome EC985 and representatives of the three ST131 clades A, B and C [15] highlighted in purple. The red box identifies the subclade with ST8196 isolates EC233 and EC234. Inset magnifies the subclade highlighted in the red box to better resolve the branching pattern and associated clade confidence scores. Clades I, II and III represent the three major clades in which the different STs within the ST131CC cluster.

EC233 and EC234 had three SNP differences, while EC70 and EC234 differed by 12 SNPs. The ST8189 representative, ESC_RA8813AA isolated in 2017, had 16 SNP differences with EC234.

The ST8196 isolates, EC233 and EC234, had the least branch distance to EC70 (ST131) indicating closest genealogical association. To explore further, we attempted a time-resolved genealogy on a global

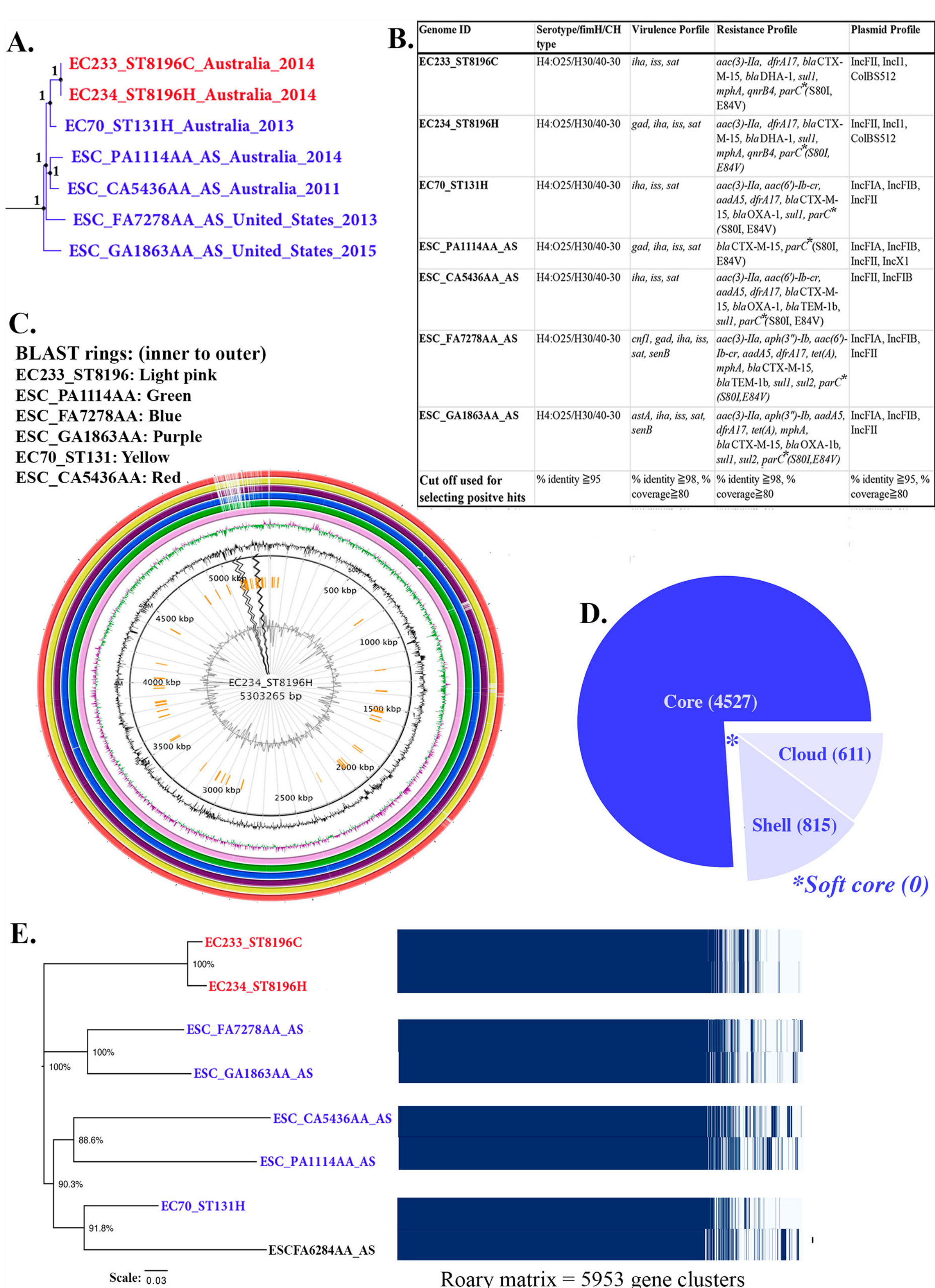
collection of ST131 genomes and the ST8196 isolates. Our objective was to identify the origin of ST8196 genomes by determining ancestral relationships with ST131 genomes circulating locally in Australia and in geographically distinct regions, which remained unidentified in the previous phylogeny analysis (Figure 1) owing to any unintended bias imposed by the selection of ST131 genomes. TempEst [42] analysis of aligned regions from 621 ST131 genomes (retrieved from the Enterobase database, 30th May 2019) included in a parSNP analysis, indicated that the coefficient of correlation between time and substitutions was 0.1 posing a challenge to select an appropriate molecular clock from the available list of models. Therefore, we constructed an unrooted phylogenetic tree with MrBayes in order to investigate the genealogy of these isolates. The ASDSF calculated from the two independent chains in MrBayes was 0.004, indicating that the chains had converged on the same posterior distribution of tree topologies. MrBayes analysis of a subset of 148 genomes (representing the widest geographical distribution for the ST131 genomes deposited in Enterobase), together with all isolates in our collection and the ST8196 genomes (File S1, S2), revealed a single subclade of seven genomes (Figure 2(A) and Figure S2) which were closely related to EC233 and EC234. The subclade comprised three isolates from Australia and two from the United States collected between 2011 and 2015 from human (blood or urine) specimens. The probability scores for resolution of subclades (displayed as node values) were 1, indicating 100% confidence in the splits. Based on branch distances between isolates in the subclade highlighted as Figure 2(A), the ST8196 representatives EC233 and EC234 were most closely related to ST131 isolate from Sydney (EC70 collected in 2013). EC70 was collected from Concord Hospital in Sydney in the course of sample collection for this study [23]. The other two Australian ST131 stains in the subclade, ESC_PA1114AA and ESC_CA5436AA were collected from Brisbane, Australia. These Australian isolates in the subclade clustered with two ST131 genomes from the USA, ESC_FA7278AA collected from urine specimen of a patient in 2013 and ESC_GA1863AA, collected in 2015 from a human blood specimen.

Genotypic profiling of ST131 and ST8196 isolates that cluster together suggest major differences are within accessory regions of the genomes

All seven isolates in the subclade (Figure 2(A)) had a O25:H4 serotype, *H*30 *fimH* type, 40–30 CH type, carried the same chromosomal *parC* (S80I, E84V) and *gyrA*(S83L, D87N) mutations (known to confer fluoroquinolone resistance) and $bla_{\text{CTX-M-15}}$ which are features that define the *H*30Rx clonal lineage of ST131 [15,16]. In addition, they share a common pool of virulence-associated genes, including *iha* (encoding adherence protein), *iss* (increased serum survival gene, supporting immune evasion) and *sat* (secreted autotransporter toxin) (Figure 2(B)). Three isolates also carry one (ESC_PA1114AA_AS, ESC_FA7278AA_AS) or two copies (EC234_ST8196) of the acid neutralization and survival gene, *gad*, encoding glutamate decarboxylase. Isolate ESC_FA7278AA_AS from the United States additionally carries *cnf1* (cytotoxic necrotising factor) and *senB* (plasmid-encoded enterotoxin), genes that are absent in the Australian isolates. A comparison of the chromosomally located efflux pumps commonly associated with drug resistance in *E. coli* also indicated identical profiles in isolates EC233_ST8196, EC234_ST8196 and EC70_ST131 (File S3A). Major differences in ST8196 and ST131 genomes were identified in the acquired resistance gene pool and plasmid profiles as presented in details in Figure 2(B), indicating contribution of laterally acquired genes in the emergence of ST8196.

A pairwise BLASTn alignment (Figure 2(C)) of the six closely related genomes with EC234_ST8196 revealed >98% sequence identity over the length of the genomes (Figure 2(C)). Major differences were noticed in sections of the BLAST rings (beyond the 5MB coordinate) where smaller contigs of the partially closed EC234 genome, comprising the four closed plasmids and the four additional contigs were present. The innermost circle (solid light pink colour) in Figure 2(C) represents an alignment of EC234 with EC233_ST8196 and revealed >99.9% identity over the entire length of the genomes, including the accessory regions. Gaps around genomic coordinates 1.3MB and 1.4MB and beyond 5MB in the yellow BLAST ring, representing EC70_ST131, indicated regions of difference with EC234_ST8196. Annotation of genes between 1.3MB and 1.4MB in EC234 indicated presence of phage associated genes, while regions beyond 5MB comprised plasmid associated genes.

Genomic islands (GI) defined on the basis of regions of differential codon usage [43] in the partially closed EC234 genome are indicated as yellow radiating lines inside the BLASTn rings in Figure 2(C). All isolates included in the pairwise BLASTn analysis, irrespective of the country of isolation, revealed near identity across GIs in EC234 indicating relative conservation of these loci. Gene content of these regions (File S3B) primarily comprises mobile element associated genes, hypothetical genes and some regulatory protein-encoding genes, except regions around genomic coordinates 1.62Mb and 1.9MB (File S3B) which houses the type 1 fimbrial operon recombinase genes, *fimB* and *fimE*, in EC234 (File S3B). The fibrillar regulatory genes control phase variation in uro-pathogenic *E. coli* and are located in a well-known variable region of the *E. coli* genome and are associated with the differential



Genome ID	Serotype/fimH/CH type	Virulence Profile	Resistance Profile	Plasmid Profile
EC233_ST8196C	H4:O25/H30/40-30	*iha, iss, sat*	*aac(3)-IIa, dfrA17, bla*CTX-M-15, *bla*DHA-1, *sul1, mphA, qnrB4, parC**(S80I, E84V)	IncFII, IncI1, ColBS512
EC234_ST8196H	H4:O25/H30/40-30	*gad, iha, iss, sat*	*aac(3)-IIa, dfrA17, bla*CTX-M-15, *bla*DHA-1, *sul1, mphA, qnrB4, parC**(S80I, E84V)	IncFII, IncI1, ColBS512
EC70_ST131H	H4:O25/H30/40-30	*iha, iss, sat*	*aac(3)-IIa, aac(6')-Ib-cr, aadA5, dfrA17, bla*CTX-M-15, *bla*OXA-1, *sul1, parC** (S80I, E84V)	IncFIA, IncFIB, IncFII
ESC_PA1114AA_AS	H4:O25/H30/40-30	*gad, iha, iss, sat*	*bla*CTX-M-15, *parC**(S80I, E84V)	IncFIA, IncFIB, IncFII, IncX1
ESC_CA5436AA_AS	H4:O25/H30/40-30	*iha, iss, sat*	*aac(3)-IIa, aac(6')-Ib-cr, aadA5, dfrA17, bla*CTX-M-15, *bla*OXA-1, *bla*TEM-1b, *sul1, parC**(S80I, E84V)	IncFII, IncFIB
ESC_FA7278AA_AS	H4:O25/H30/40-30	*cnf1, gad, iha, iss, sat, senB*	*aac(3)-IIa, aph(3")-Ib, aac(6')-Ib-cr, aadA5, dfrA17, tet(A), mphA, bla*CTX-M-15, *bla*TEM-1b, *sul1, sul2, parC** (S80I,E84V)	IncFIA, IncFIB, IncFII
ESC_GA1863AA_AS	H4:O25/H30/40-30	*astA, iha, iss, sat, senB*	*aac(3)-IIa, aph(3")-Ib, aadA5, dfrA17, tet(A), mphA, bla*CTX-M-15, *bla*OXA-1b, *sul1, sul2, parC**(S80I,E84V)	IncFIA, IncFIB, IncFII
Cut off used for selecting positve hits	% identity ≧95	% identity ≧98, % coverage≧80	% identity ≧98, % coverage≧80	% identity ≧95, % coverage≧80

Figure 2. Pangenome analysis of isolates that clustered closely in a MrBayes analysis (Files S1 and S2). A. Magnified view of the subclade containing ST8196 isolates and other closely related ST131 genomes identified in MrBayes analysis presented in supplementary file S2. B. Genotypic feature that characterize isolates in the subclade presented in section A. Except *parC* which is highlighted with an asterisk (*), all genes listed in the "Resistance genes" column are known to be laterally acquired. C. BLASTn alignment (using 90% lower and 100% upper identity cutoffs) of 6 genomes against the partially completed EC234_ST8196 genome. The inner radiating bars indicate the location of genomic islands in EC234, as predicted by the SIGI-HMM module of islandviewer software. D. Graphical overview of distribution of genes that represent the core and accessory genomes of the isolates in Roary. (Core genes = genes present in 99% <= strains < = 100%), (Soft core genes = genes present in 95% <= strains < 99%, shell genes = genes present in 15% <= strains < 95%), Cloud genes (genes present in 0% <= strains < 15%). E. Genome clustering and the pangenome matrix of 8 genomes (including the outgroup, ESCFA6284_FA). Numbers at the node of the tree indicate bootstrap values.

transcriptional regulation of pathogenic factors in specific *E. coli* strains [44].

Pangenome analysis of ST131 subclade with ST8196

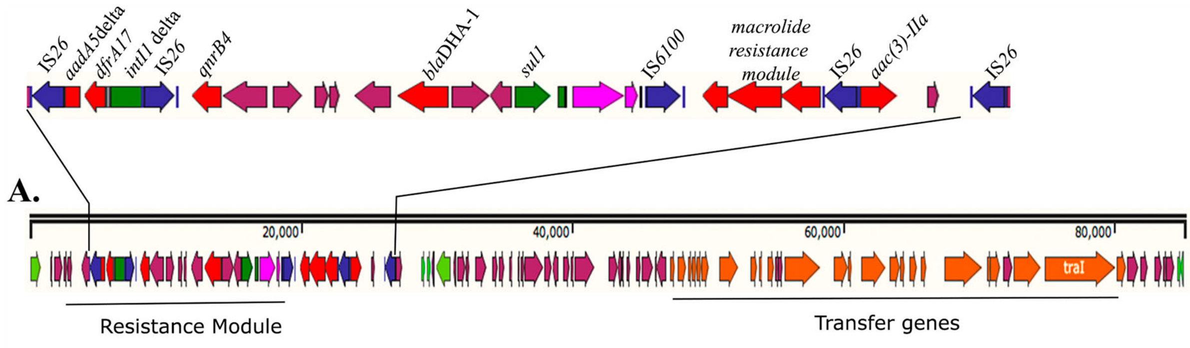
Roary [34] was used to map the core-genome and accessory genome content of the seven closely related isolates. ESCFA6284AA, collected from a human blood sample in the USA in 2015 was included as an outgroup. Based on ≥95% amino acid sequence identity and MCL clustering (Markov clustering) of peptides encoded by open reading frames, 5953 clusters constituted the pangenome of all eight isolates, of which 4527 (Figure 2(D)) clusters comprised the hard-core of the genomes. In pan-genome analysis, core genome content is often divided into hard-core and soft-core, with hard-core gene clusters expected to be present in all gnomes while gene clusters constituting soft-core present in ≥95% genomes. The accessory genome content was divided into two subsets, the "shell" (815) and the "cloud" (611). "Shell" genes represent a pool of shared accessory genes present in the majority of genomes analysed, while "cloud" genes comprise unique clusters present in at least 1 genome in the cohort. Our analysis indicated that 4663 gene clusters were present in at least seven of the eight genomes (File S4), while the variable pool of accessory genes comprised 611 clusters. Isolate ESC_-CA5436AA, clearly had the most variable subset of unique gene clusters (Figure 2(E), File S4). Clustering of genomes based on the presence and absence of accessory genes (Figure 2(E)) indicate that major differences between ST8196 (EC234 and EC233) and ST131 genomes were in the accessory gene content. Combining datasets generated in pangenome analysis (based on amino acid identity of predicted ORFs) with whole-genome SNP based phylogeny analyses, it is evident that major differences between Australian ST131-*H*30Rx subclade and the ST8196 isolates, EC234 and EC233 (Figure 2(E,A)), were in the accessory gene pool.

Gene clusters in the accessory genome pool primarily consisted of plasmid associated genes, resistance genes and hypotheticals. Co-relating pan-genome analysis outputs with genotype profiling data, EC233 and EC234 had three typeable plasmid replicons, an IncFII plasmid, an IncI1 plasmid and a ColBS512-like plasmid. The IncI1 and ColBS512-like plasmid replicons were unique to the two ST8196 isolates and were not identified in any of the closely related ST131 genomes selected from MrBayes analysis. However, ColBS512-like plasmid has been identified in other ST131 isolates [33], and the replicon was identified in FA6284AA_AS_ST131 the outgroup included in our Roary analysis (Figure 2(B)). All 5 closely related ST131 genomes selected via the MrBayes analysis had IncFIA, IncFII and IncFIB replicons. Isolate PA1114AA_AS, an ST131 isolate collected from BSI of a patient in Brisbane, Australia additionally carried an IncX1 plasmid.

Plasmids in EC234_ST8196

Unicycler hybrid assembly of long and short-read sequences of EC234 resolved the genome into nine contigs, including a 5.09Mb contig representing a major portion of the chromosome, 4 closed plasmids, and 4 additional contigs which likely form parts of the chromosome. Contig 2 typed as a 92,955nt IncI1 (pSPRC_Ec234-I) plasmid, while contig 3 represented a 85,199nt long IncFII_2 (pSPRC_Ec234-FII) plasmid. The isolate also had two small plasmids. One of these, pSPRC_Ec234-3 (contig 5; 5164nt), only housed *repA* and *mobA* genes in addition to 3 hypothetical genes. pSPRC_Ec234-3 displayed 99.7% identity over 100% of the query sequence with pEc631_5 (CP040268.1) isolated from an *E. coli* inhabiting a marine bivalve mollusc in Norway. There are several other *E. coli* plasmids in GenBank which share ≥99% identity over the length of pSPRC_Ec234-3 collected from geographically disparate regions. The second small plasmid pSPRC_Ec234-4, represented by contig 6 (112.8x coverage), was 2,101nt long and was identical to a *Shigella sonnei* Ss046 (pSS046_spC, CP000643.1) plasmid isolated from a patient with bacillary dysentery in China. Plasmid finder typed it as a ColBS512-like plasmid, as the *rep* gene was identical to *S. boydii* CDC 3083-94 plasmid pBS512_2 (CP001058.1). Pairwise alignment of pSPRC_Ec234-4 and pBS512_2 displayed no sequence similarity across the remaining length of pBS512_2.

Apart from $bla_{\text{CTX-M-15}}$ that was located on the chromosome of EC234 (contig 1) adjacent to an IS*Ecp1* insertion element and *gyrB* and *parC* mutations that contribute to fluoroquinolone resistance, the IncFII_2 plasmid pSPRC_Ec234-FII carried most of the genes that accounted for the phenotypic resistance profile (Figure 3(A)) of the isolate. Plasmid pSPRC_Ec234-FII also carried *qnrB4* and $bla_{\text{DHA-1}}$ in close proximity to the CRL. There were four copies of IS*26* in the resistance locus. Two direct copies of IS*26* flanked the locus indicating that it can potentially function as a translocatable unit. The locus contains diverse resistance gene cargo including a fragmented class 1 integron with a *dfrA17-aadA5* gene cassette-array. Notably, both the integrase (*intI1*) and the aminoglycoside resistance (*aadA5*) genes were interrupted by copies of IS*26*. An additional copy of IS*26* resides between the IS*6100* associated macrolide resistance module and the aminoglycoside resistance gene *aac (3)-IIa* (encoding resistance to gentamicin and tobramycin). Other features of the resistance loci included remnants of the 3′-CS of a class 1 integron and the



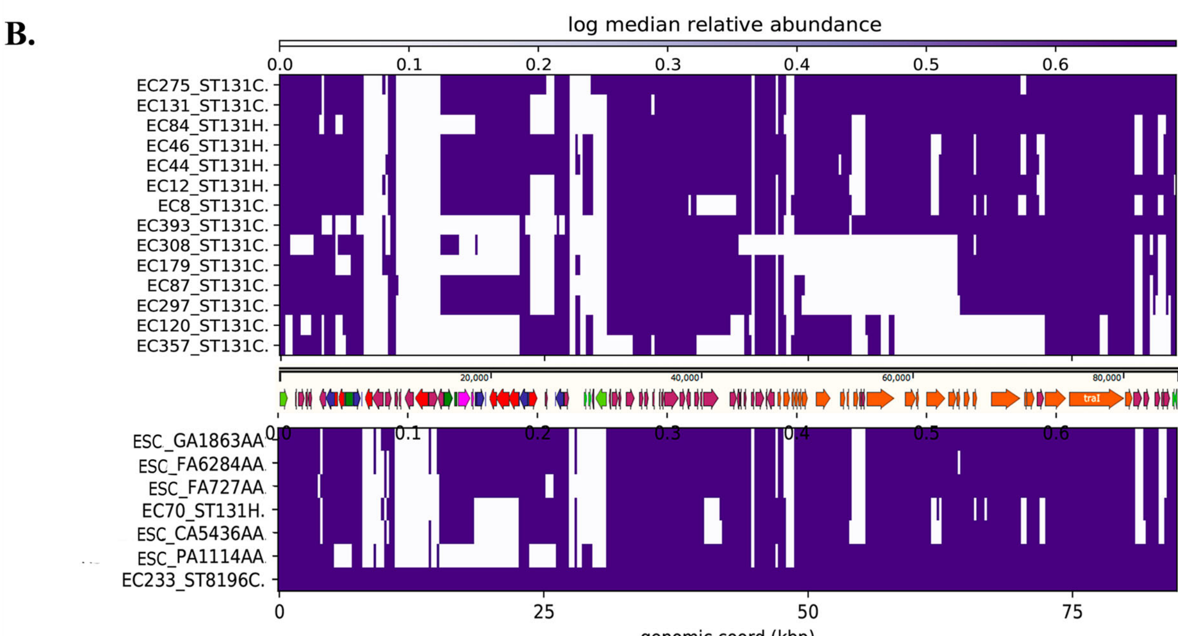


Figure 3. pSPRC_Ec234-FII the IncFII plasmid in EC234. A. Diagrammatic representation of gene content in the IncFII plasmid of isolate EC234 and the resistance loci in the inset above. Arrows representing gene families are colour coded. Orange arrows: genes required for plasmid transfer, light green: replication genes, dark green: class 1 integron integrase gene, red arrows: resistance genes, dark blue: insertion elements and purple: hypothetical genes. B. Mapping of raw illumina reads from ST131 in our collection (upper panel) and EC233_ST8196 and other ST131 isolates identified to be closely related to EC234 from Mr Bayes phylogeny analysis (bottom panel).

bla$_{DHA-1}$ beta-lactamase gene. BLASTn analysis of pSPRC_Ec234-FII identified pUB-DHA-1 (MK048477.1) from *E. coli* strain U-5227, isolated from human urine, as the closest match in GenBank. pSPRC_Ec234-FII shared 99.93% identity over 96% length of pUB-DHA-1, including the cargo of resistance genes. The presence of four copies of IS*26* may enable multiple sections of the resistance locus to independently mobilize as translocatable units containing different combinations of resistance genes [45,46].

IncFII replicons are common features of ST131 genomes [2,15,16,40]. Most phylogenetically related ST131 isolates identified in this study tested positive for IncFII replicons and had identical resistance cargoes (Figure 2 (B)). We, therefore, verified the presence of pSPRC_Ec234-FII within closely related ST131 genomes identified in our MrBayes phylogeny analysis (Figure 3(B)) and in ST131 isolates in our extended collection (Hastak et al Microbial Genomics, 2020). Only EC233 (the ST8196 representative) had an identical plasmid. Our read mapping approach indicated the presence of parts of the plasmid backbone, however, the plasmid partitioning and stability genes (represented with light green arrows in the plasmid map Figure 3(B)) in pSPRC_Ec234-FII are missing in the other genomes. As such ST8196 isolates carry an IncFII_2 plasmid, variants of which likely distinguish them from the population of IncFII plasmids circulating within *E. coli* ST131 in Australia. The resistance cargo present on pSPRC_Ec234-FII has recently been reported on IncFIA and IncI1 plasmids carried by ST131 in Australia [47], indicating possible movement of a compound resistance transposon. pSPRC_Ec234-FII has a full complement of transfer genes and is likely laterally mobile, although plasmid mobility was not tested in this study.

The IncI1 plasmid in EC234, pSPRC_Ec234-I, did not carry any virulence or resistance genes, however, it had the full complement of conjugative (type IV pili) and transfer genes, indicating an ability to

mobilize independently and perhaps co-mobilize other resident plasmids. A BLASTn analysis of pSPRC_Ec234-I identified pSTM7 (KF290377.1) from a case of human Salmonellosis linked to pathogenic *Salmonella* Typhimurium strain STm7 in Australia [48] as the closest match in GenBank. pSPRC_Ec234-I had 99.4% identity but across 73% of the length of pSTM7. To examine the relative prevalence of IncI1 plasmids similar to pSPRC_Ec234-I in the plasmid database, 500 completely closed IncI plasmids were downloaded (23rd September 2019) [49]. A subset of 38 IncI1 plasmids sourced from *E. coli, Shigella* and *Salmonella* spp. that showed >99% identity over >60% of the length of pSPRC_Ec234-I, were selected for SNP phylogeny analysis (Figure 4(A)). Based on the alignment of ≥59% of the length of pSPRC_Ec234-I with the 38 plasmids selected, pSPRC_Ec234-I grouped with three other IncI1 plasmids, pU12A_B (NZ_CP035469.1), pU15A_A (NZ_CP035721.1) and an un-named *E. coli* plasmid (NZ_LS992187.1). While pU12A_B and pU15A_A were sourced from *E. coli* ST131 isolates recovered from urine samples in Australia, the *E. coli* plasmid in NZ_LS992187.1 was reported from Germany. Considering branch lengths in the SNP tree, pSPRC_Ec234-I is most closely related to the plasmid in the German isolate (NZ_LS992187.1), while the two Australian plasmids NZ_CP035469.1 and NZ_CP035721.1 [47] appeared to have accumulated more SNPS and indels. Alignment of all three plasmids revealed that the main differences were in the hypothetical genes, while the backbone was fairly conserved, further supporting the hypothesis that closely related IncI1 plasmids are circulating in Australia.

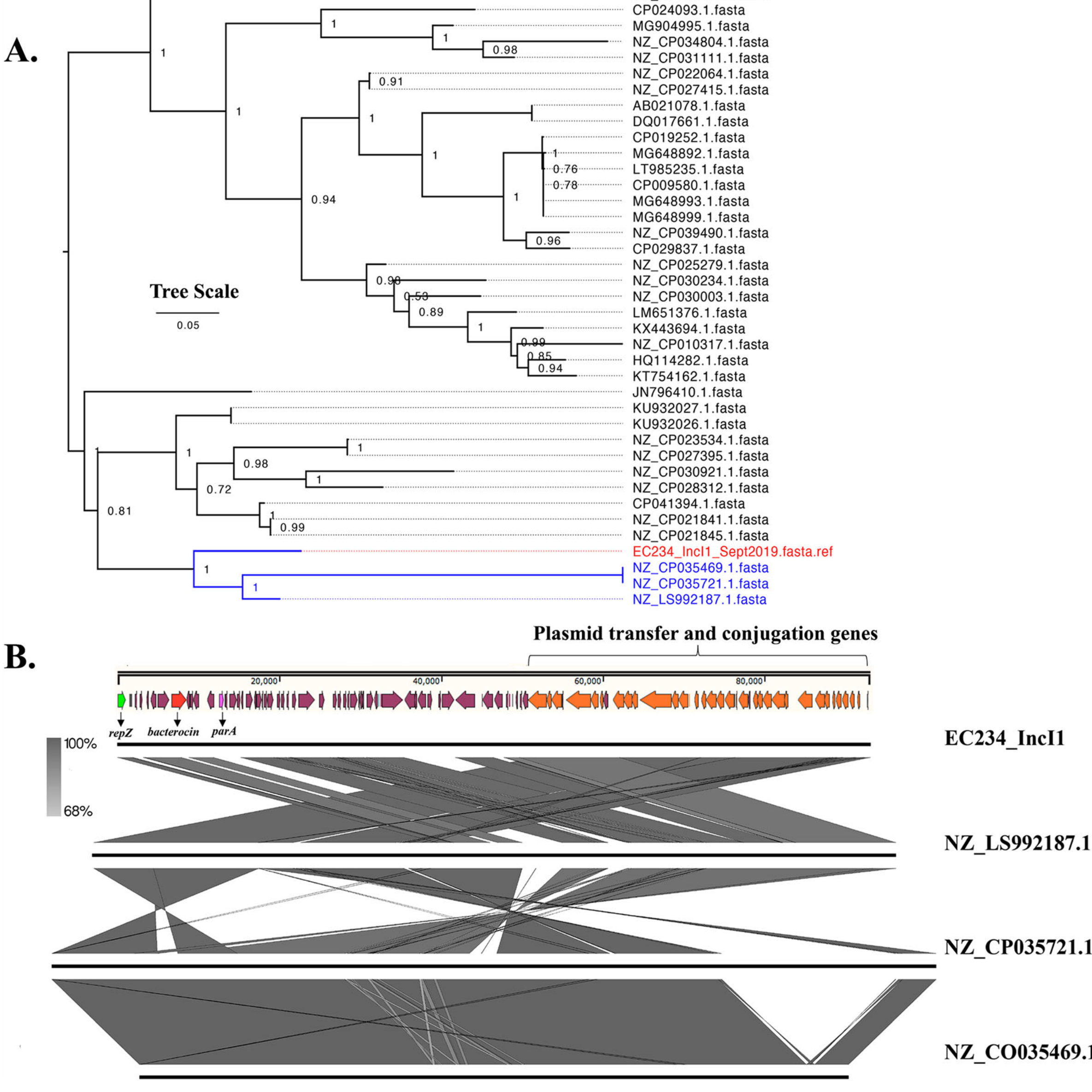


Figure 4. Features of IncI1 plasmid pSPRC_Ec234-I in strain EC234. A. Phylogeny of complete IncI1 plasmids available in the plasmid database. Orange arrows indicate genes required for plasmid transfer and conjugation, green indicates replication genes, red arrows indicate bacteriocin resistance genes and purple arrows indicate hypothetical genes. B. EasyFig alignment of the three most closely related plasmids identified in isolate EC234.

Discussion

Pandemic and high resistance-risk clonal lineages of extra-intestinal *E. coli* cause serious problems for human health globally. We present the first description of a novel, extensively drug-resistant ExPEC lineage, ST8196, within the ST131 clonal complex. The two representative isolates, EC234 and EC233, were collected one day apart from different patients presenting with bacteraemia at a hospital in Sydney in November 2014 and are characterized by whole genome sequence analysis in this study. Although the 48 hr post admission cut-off used to classify hospital-onset and community-onset infections can be misleading, the isolates were collected from two different patients indicating the presence of the strain in the local community serviced by the hospital. Our phylogenetic analyses provide compelling evidence that these clonal isolates, and the ST8196 lineage have most likely evolved in Australia from the *H*30Rx sub-lineage of *E. coli* ST131 that are circulating locally. Major differences in ST131 and ST8196 were found primarily in the accessory genomes. In EC234, the accessory genome included four completely closed plasmids, an IncFII_2 plasmid pSPRC_Ec234-FII (85,199bp), two small plasmids pSPRC_Ec234-3 (5,164bp) and pSPRC_Ec234-4 (2,101bp) and a novel IncI1 plasmid pSPRC_Ec234-I (92,955bp). The IncFII_2 plasmid pSPRC_Ec234-FII carries the majority of the resistance gene cargo contributing towards the phenotypic resistance profile and is likely a variant of IncFII plasmids circulating within ST131-cladeC2 in Australia [33]. The IncI1 plasmid pSPRC_Ec234-I in EC234 is likely a progenitor of IncI1 plasmid lineages reported from ST131 in Australia [47]. The two IncI1 plasmids reported recently from Australia pU12A_B (NZ_CP035469.1) and pU15A_A (NZ_CP035721.1) [47] may have evolved from a common ancestral IncI1 lineage, raising the possibility that ST8196 strains EC233 and EC234 acquired pSPRC_Ec234-I, or a close variant of it, locally.

The dissemination of MDR plasmids within clinical settings often complicates treatment options. Plasmids belonging to IncFII incompatibility groups and their variants are well established in ST131 populations globally [2, 33]. These plasmids frequently carry allelic variants of $bla_{CTX\text{-}M}$ gene encoding extended spectrum β-lactamase (ESBL) enzymes that hydrolyse cephalosporins and aztreonam, and co-transfer plasmid-mediated quinolone and aminoglycoside resistance [50,51]. The CTX-M family of ESBL enzymes are classified into different phylogenetic groups [52] based on amino acid sequence of the peptides encoded by the respective allelic variant. The CTX-M cluster 1 genes, which include CTX-M-15 allele, is primarily disseminated via IncFII:IncFIA plasmids and sometimes on IncI plasmids [19]; while the CTX-M cluster 9 genes, including CTX-M-14 and CTX-M-27 are usually associated with IncHI2 plasmids and sometimes with IncFII [53,54]. In the ST8196 isolates, the $bla_{CTX\text{-}M\text{-}15}$ gene was not associated with any plasmid.

To identify and establish the location of resistance genes in the genomes, including the $bla_{CTX\text{-}M\text{-}15}$ gene, we assembled EC234 genome using both long read and short-read sequences. The hybrid assembly of EC234 resolved four plasmids. All resistance genes contributing to the resistance profile were found to cluster within a complex resistance locus flanked by direct copies of IS*26* on IncFII_2 plasmid pSPRC_Ec234-FII. The IS*26* flanked region contained a *qnrB4* quinolone resistance gene and a $bla_{DHA\text{-}1}$ β-lactamase gene in close proximity to the CRL. The *qnrB4* and $bla_{DHA\text{-}1}$ genes are often linked together and responsible for the dispersal of plasmid-mediated quinolone resistance in Enterobacteriaceae [50,55,56]. Genes conferring resistance to extended-spectrum β-lactam ($bla_{CTX\text{-}M\text{-}15}$) and fluoroquinolone (*gyrA1*AB and *parC1a*AB) antibiotics were located on the largest contig in *E. coli* ST8196; with $bla_{CTX\text{-}M\text{-}15}$ linked to insertion element IS*Ecp1*. In addition to providing lateral mobility, IS*EcpI* also provides a strong promoter for expression of the $bla_{CTX\text{-}M\text{-}15}$ gene [54,57]. The 2971bp IS*Ecp1*-$bla_{CTX\text{-}M15}$ module in EC234 is identical to a module found on various plasmid backbones in several entries in the GenBank database, however, in EC234 it is inserted at a unique intergenic region of the chromosome. The module is flanked by 5bp (TGATC) direct repeats, indicating a single transposition event associated with the genomic location. Although the chromosome in EC234 was not closed and may have some unresolved chromosomal configurations, genes contributing to fluoroquinolone resistance and ESBL production were not located near scaffold breaks and were identified at identical locations in previous studies [2,14]. Chromosomal incorporation of an IS*Ecp1* associated $bla_{CTX\text{-}M\text{-}15}$ gene within fluoroquinolone-resistant *E. coli* ST131 isolates has previously been reported in members of the *H*30Rx subclade [13,15], but the insertion site identified in the ST8196 isolates is unique.

A recent pangenome analysis on a global collection of 4071 ST131 isolates predicts the potential of outbreaks by genetically variable isolates with altered resistance gene cargoes [14] and highlights the necessity for constant genomic surveillance. Different STs that cluster together into clonal complexes (CC) comprise such genetically similar isolates which are characterized by a single change in the allelic profile of the target genes defining the central MLST genotype [20]. While ST131 continues to be the focus of many genomic and epidemiological studies, there is an underappreciation of other STs within the ST131CC. Our data highlights the near identity of ST8196 isolates with strains in ST131-*H*30Rx subclade. The acquisition of IncI1 and the small cryptic plasmids by ST8196

identified in the course of this study appears to have contributed to the evolution of this ST within the ST131CC. Our study also unravels the phylogenetic relationship of different STs within the "ST131-clonal complex" which cluster with representatives of the three well known ST131 clades; and identifies 43 other STs within ST131CC that are closely related to the ST131-clade C representatives. They are therefore worth exploring further in terms of phenotype-genotype correlations and monitoring in the course of genomic epidemiological studies. ST8196 also appears to be closely related to *E. coli* ST8189, identified recently in Australia, however, EC233 and EC234 share a more recent common ancestor with *E. coli* strain EC70_ST131 isolated from the same Sydney hospital. Based on the MLST profiles, ST8189 differed from ST131 in the sequence of *mdh* (allelic variant number 697) encoding malate dehydrogenase. Metadata pertaining to the country from which the ST8196 was isolated was not available in Enterobase or online. However, it is noteworthy that currently there are three other representatives of ST8189 in Enterobase, including an isolate (ESC_OA1826AA, SRR10126950) collected in 2017 from Australia and are worth investigating as they are also phylogenetically related to the *H*30Rx subclade.

The proposed evolution of ST8189 is based on an extensive phylogenetic analysis of the core genome and is supported by pangenome analysis. Our data suggests that the ST8196 isolates described here evolved from the *H*30Rx subclade of ST131 and are armed with virulence attributes and antimicrobial resistance gene cargo that may contribute to the local establishment of these clones. In addition, the presence of multiple copies of IS*26* within the resistance locus suggests that new genetic cargo mobilized by IS*26* can be readily acquired, and existing resistance cargo can be passed to other lineages within the Enterobacteriales. ST8196 isolates have the potential to cause disease with significantly curtailed treatment options. We identify 44 STs within the ST131CC that are closely related to the globally dominant ST131-Clade C strains, highlighting the necessity of improved characterization of these clinically important *E. coli* lineages. Our study emphasizes the importance of unbiased genomic surveillance for detecting novel, MDR ExPEC circulating in the community and which may ultimately present at health care facilities.

Acknowledgements

PH is a recipient of the University of Technology Sydney PhD scholarship as a part of the AusGEM programme and the University of Technology International Research Scholarship (IRS).

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This project was funded by the Australian Centre for Genomic Epidemiological Microbiology (AusGEM), a strategic partnership between the NSW-Department of Primary Industries and the University of Technology Sydney. The results presented in this manuscript are helping to inform the OUTBREAK decision support system funded by the Medical Research Future Fund Frontier Health and Medical Research Program (grant number MRFF75873).

ORCID

Piklu Roy Chowdhury http://orcid.org/0000-0003-4352-603X

References

[1] Pitout JD. Extraintestinal pathogenic Escherichia coli: a combination of virulence with antibiotic resistance. Front Microbiol. 2012;3:9.
[2] Pitout JD, DeVinney R. Escherichia coli ST131: a multidrug-resistant clone primed for global domination. 2017;6:F1000 Faculty Rev-195.
[3] Cho S, Gupta SK, McMillan EA, et al. Genomic analysis of multidrug-resistant Escherichia coli from surface water in Northeast Georgia, United States: presence of an ST131 epidemic strain containing blaCTX-M-15 on a phage-like plasmid. Microb Drug Resist. 2019 Nov 14;26(5):447–455.
[4] Cho S, Nguyen HAT, McDonald JM, et al. Genetic characterization of antimicrobial-resistant Escherichia coli isolated from a mixed-use watershed in Northeast Georgia, USA. Int J Environ Res Public Health. 2019 Oct 7;16(19):3761.
[5] Melo LC, Haenni M, Saras E, et al. Emergence of the C1-M27 cluster in ST131 Escherichia coli from companion animals in France. J Antimicrob Chemother. 2019 Oct 1;74(10):3111–3113.
[6] Reid CJ, McKinnon J, Djordjevic SP. Clonal ST131-H22 Escherichia coli strains from a healthy pig and a human urinary tract infection carry highly similar resistance and virulence plasmids. Microb Genom. 2019 Sep;5(9):e000295.
[7] Seenama C, Thamlikitkul V, Ratthawongjirakul P. Multilocus sequence typing and bla ESBL characterization of extended-spectrum beta-lactamase-producing Escherichia coli isolated from healthy humans and swine in Northern Thailand. Infect Drug Resist. 2019;12:2201–2214.
[8] Banerjee R, Johnson JR. A new clone sweeps clean: the enigmatic emergence of Escherichia coli sequence type 131. Antimicrob Agents Chemother. 2014 Sep;58 (9):4997–5004.
[9] Burgess MJ, Johnson JR, Porter SB, et al. Long-term care facilities are reservoirs for antimicrobial-resistant sequence type 131 Escherichia coli. Open Forum Infect Dis. 2015 Jan;2(1):ofv011.
[10] Diekema DJ, Hsueh PR, Mendes RE, et al. The microbiology of bloodstream Infection: 20-year trends from the SENTRY antimicrobial surveillance program. Antimicrob Agents Chemother. 2019 Jul;63(7): e00355–19.
[11] Mineau S, Kozak R, Kissoon M, et al. Emerging antimicrobial resistance among Escherichia coli strains in bloodstream infections in Toronto, 2006-2016: a

retrospective cohort study. CMAJ Open. 2018 Oct-Dec;6(4):E580–E586.

[12] Poolman JT, Wacker M. Extraintestinal pathogenic Escherichia coli, a common human pathogen: challenges for vaccine development and progress in the field. J Infect Dis. 2016 Jan 1;213(1):6–13.

[13] Ben Zakour NL, Alsheikh-Hussain AS, Ashcroft MM, et al. Sequential acquisition of virulence and fluoroquinolone resistance has shaped the evolution of Escherichia coli ST131. mBio. 2016 Apr 26;7(2):e00347–16.

[14] Decano AG, Downing T. An Escherichia coli ST131 pangenome atlas reveals population structure and evolution across 4,071 isolates. Sci Rep. 2019 Nov 22;9(1):17394.

[15] Petty NK, Ben Zakour NL, Stanton-Cook M, et al. Global dissemination of a multidrug resistant Escherichia coli clone. Proc Natl Acad Sci U S A. 2014 Apr 15;111(15):5694–5699.

[16] Price LB, Johnson JR, Aziz M, et al. The epidemic of extended-spectrum-beta-lactamase-producing Escherichia coli ST131 is driven by a single highly pathogenic subclone, H30-Rx. MBio. 2013 Dec 17;4(6):e00377–13.

[17] Stoesser N, Sheppard AE, Pankhurst L, et al. Evolutionary history of the global emergence of the Escherichia coli epidemic Clone ST131. mBio. 2016 Mar 22;7(2):e02162.

[18] Mathers AJ, Peirano G, Pitout JD. Escherichia coli ST131: the quintessential example of an international multiresistant high-risk clone. Adv Appl Microbiol. 2015;90:109–154.

[19] Johnson TJ, Danzeisen JL, Youmans B, et al. Separate F-type plasmids have shaped the evolution of the H30 Subclone of Escherichia coli sequence type 131. mSphere. 2016 Jul-Aug;1(4):e00121–16.

[20] Feil EJ, Li BC, Aanensen DM, et al. eBURST: inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus sequence typing data. J Bacteriol. 2004 Mar;186(5):1518–1530.

[21] Rasko DA, Rosovitz MJ, Myers GS, et al. The pangenome structure of Escherichia coli: comparative genomic analysis of E. coli commensal and pathogenic isolates. J Bacteriol. 2008 Oct;190(20):6881–6893.

[22] Dahbi G, Mora A, Lopez C, et al. Emergence of new variants of ST131 clonal group among extraintestinal pathogenic Escherichia coli producing extended-spectrum beta-lactamases. Int J Antimicrob Agents. 2013 Oct;42(4):347–351.

[23] Hastak P, Cummins ML, Gottlieb T, et al. Genomic profiling of Escherichia coli isolates from bacteraemia patients: a 3-year cohort study of isolates collected at a Sydney teaching hospital. Microb Genom. 2020 May 6;6(5):e000371.

[24] Magiorakos AP, Srinivasan A, Carey RB, et al. Multidrug-resistant, extensively drug-resistant and pandrug-resistant bacteria: an international expert proposal for interim standard definitions for acquired resistance. Clin Microbiol Infect. 2012 Mar;18(3):268–281.

[25] Bell JM, Turnidge JD, Coombs GW, et al. Australian Group on antimicrobial resistance Australian Enterobacteriaceae Sepsis Outcome Programme, Annual Report 2014. In: Curran M, editor. Communicable diseases intelligence. June 2016 ed. Australia: The Department of Health; 2016. p. 229–235.

[26] Darling AE, McKinnon J, Worden P, et al. A draft genome of Escherichia coli sequence type 127 strain 2009-46. Gut Pathog. 2014;6:32.

[27] Wick RR, Judd LM, Gorrie CL, et al. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. PLoS Comput Biol. 2017 Jun;13(6):e1005595.

[28] Brettin T, Davis JJ, Disz T, et al. RASTtk: a modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. Sci Rep. 2015 Feb 10;5:8365.

[29] Altschul SF, Madden TL, Schaffer AA, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 1997 Sep 1;25(17):3389–3402.

[30] Gertz EM, Yu YK, Agarwala R, et al. Composition-based statistics and translated nucleotide searches: improving the TBLASTN module of BLAST. BMC Biol. 2006 Dec 7:41.

[31] Schaffer AA, Aravind L, Madden TL, et al. Improving the accuracy of PSI-BLAST protein database searches with composition-based statistics and other refinements. Nucleic Acids Res. 2001 Jul 15;29(14):2994–3005.

[32] Alikhan NF, Petty NK, Ben Zakour NL, et al. BLAST ring Image Generator (BRIG): simple prokaryote genome comparisons. BMC Genomics. 2011 Aug 8;12:402.

[33] Sullivan MJ, Petty NK, Beatson SA. Easyfig: a genome comparison visualizer. Bioinformatics. 2011 Apr 1;27(7):1009–1010.

[34] Page AJ, Cummins CA, Hunt M, et al. Roary: rapid large-scale prokaryote pan genome analysis. Bioinformatics. 2015 Nov 15;31(22):3691–3693.

[35] R Core Team. R: a language and environment for stastical computing. R Foundation for Statistical computing. Vienna: European Environmental Agency; 2019.

[36] Treangen TJ, Ondov BD, Koren S, et al. The harvest suite for rapid core-genome alignment and visualization of thousands of intraspecific microbial genomes. Genome Biol. 2014;15(11):524.

[37] Forde BM, Ben Zakour NL, Stanton-Cook M, et al. The complete genome sequence of Escherichia coli EC958: a high quality reference sequence for the globally disseminated multidrug resistant E. coli O25b:H4-ST131 clone. PLoS One. 2014;9(8):e104400.

[38] Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014 May 1;30(9):1312–1313.

[39] Ronquist F, Teslenko M, van der Mark P, et al. Mrbayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst Biol. 2012 May;61(3):539–542.

[40] Letunic I, Bork P. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic Acids Res. 2016 Jul 8;44(W1):W242–W245.

[41] McNally A, Oren Y, Kelly D, et al. Combined analysis of variation in core, accessory and regulatory genome regions provides a super-resolution view into the evolution of bacterial populations. PLoS Genet. 2016 Sep;12(9):e1006280.

[42] Rambaut A, Lam TT, Max Carvalho L, et al. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). Virus Evol. 2016 Jan;2(1):vew007.

[43] Waack S, Keller O, Asper R, et al. Score-based prediction of genomic islands in prokaryotic genomes using hidden Markov models. BMC Bioinformatics. 2006 Mar 16;7(142).

[44] Holden N, Totsika M, Dixon L, et al. Regulation of P-fimbrial phase variation frequencies in Escherichia coli CFT073. Infect Immun. 2007 Jul;75(7):3325–3334.

[45] Harmer CJ, Hall RM. IS26-mediated precise excision of the IS26-aphA1a translocatable unit. MBio. 2015 Dec 8;6(6):e01866–15.

[46] Harmer CJ, Moran RA, Hall RM. Movement of IS26-associated antibiotic resistance genes occurs via a translocatable unit that includes a single IS26 and preferentially inserts adjacent to another IS26. MBio. 2014 Oct 7;5(5):e01801–14.

[47] Forde BM, Roberts LW, Phan MD, et al. Population dynamics of an Escherichia coli ST131 lineage during recurrent urinary tract infection. Nat Commun. 2019 Aug 13;10(1):3643.

[48] Hawkey J, Edwards DJ, Dimovski K, et al. Evidence of microevolution of Salmonella Typhimurium during a series of egg-associated outbreaks linked to a single chicken farm. BMC Genomics. 2013 Nov 19;14:800.

[49] Brooks L, Kaze M, Sistrom M. A curated, comprehensive database of plasmid sequences. Microbiol Resour Announc. 2019 Jan;8(1):e01325–18.

[50] Yang H, Chen H, Yang Q, et al. High prevalence of plasmid-mediated quinolone resistance genes qnr and aac(6')-Ib-cr in clinical isolates of Enterobacteriaceae from nine teaching hospitals in China. Antimicrob Agents Chemother. 2008 Dec;52(12):4268–4273.

[51] Yang L, Yang L, Lu DH, et al. Co-prevalance of PMQR and 16S rRNA methylase genes in clinical Escherichia coli isolates with high diversity of CTX-M from diseased farmed pigeons. Vet Microbiol. 2015 Aug 5;178 (3-4):238–245.

[52] Bonnet R. Growing group of extended-spectrum beta-lactamases: the CTX-M enzymes. Antimicrob Agents Chemother. 2004 Jan;48(1):1–14.

[53] Bevan ER, Jones AM, Hawkey PM. Global epidemiology of CTX-M beta-lactamases: temporal and geographical shifts in genotype. J Antimicrob Chemother. 2017 Aug 1;72(8):2145–2155.

[54] Poirel L, Bonnin RA, Nordmann P. Genetic support and diversity of acquired extended-spectrum beta-lactamases in Gram-negative rods. Infect Genet Evol. 2012 Jul;12(5):883–893.

[55] Robicsek A, Strahilevitz J, Sahm DF, et al. qnr prevalence in ceftazidime-resistant Enterobacteriaceae isolates from the United States. Antimicrob Agents Chemother. 2006 Aug;50(8):2872–2874.

[56] Verdet C, Benzerara Y, Gautier V, et al. Emergence of DHA-1-producing Klebsiella spp. in the Parisian region: genetic organization of the ampC and ampR genes originating from Morganella morganii. Antimicrob Agents Chemother. 2006 Feb;50(2):607–617.

[57] Poirel L, Gniadkowski M, Nordmann P. Biochemical analysis of the ceftazidime-hydrolysing extended-spectrum beta-lactamase CTX-M-15 and of its structurally related beta-lactamase CTX-M-3. J Antimicrob Chemother. 2002 Dec;50(6):1031–1034.