

Emergence of a carbapenem-resistant atypical uropathogenic *Escherichia coli* clone as an increasing cause of urinary tract infection

Received: 3 June 2025

Accepted: 21 August 2025

Published online: 02 September 2025



Lachlan L. Walker^{1,2}, Minh-Duy Phan^{1,2}, Budi Permana^{1,2}, Zheng Jie Lian^{1,2}, Nguyen Thi Khanh Nhu^{1,2}, Thom Cuddihy^{1,2}, Kate M. Peters^{1,2}, Kay A. Ramsay³, Chelsea Stewart^{1,2}, Niels Pfennigwerth⁴, Timothy J. Kidd⁵, Patrick N. A. Harris^{3,5}, David L. Paterson⁶, Matthew J. Sweet^{1,2}, Brian M. Forde^{1,2,3}✉ & Mark A. Schembri^{1,2,7}✉

Carbapenem-resistant *Enterobacterales* pose a critical global health threat, exemplified by increasing resistance of uropathogenic *E. coli* (UPEC) that cause urinary tract infections (UTIs). Here, we investigate the publicly available EnteroBase dataset and identify a signal of increasing UTI caused by phylogroup A *E. coli* sequence type 167 (ST167). Phylogenetic analysis of ST167 based on whole genome sequence data reveal three major clades (A, B, C), with clade C further resolving into several subclades, notably subclade C2 that possessed high carriage rates of carbapenem and cephalosporin resistance genes. Hierarchical clustering of core genome multi-locus sequence typing reveals ~77% of subclade C2 strains contain <20 allelic differences in 2,513 core genes and harbour two distinct group 1 capsule types, KL124 and KL30, likely originating from *Klebsiella*. Subclade C2 was predominantly sequenced in North America, and we provide evidence for clonal expansion since 2016. Analysis of UPEC virulence factors reveals complete loss of the type 1 fimbriae genes in strains from clades B and C. Two subclade C2 isolates exhibit significantly reduced capacity to colonise the bladder compared to the reference UPEC strain CFT073 in a murine UTI model. Collectively, our data identifies ST167 as an atypical UPEC clone associated with UTI.

Urinary tract infections (UTIs) are one of the most common bacterial infections worldwide, with an estimated ~404 million cases per year and an annual economic burden of US\$3.5 billion in the United States alone^{1,2}. UTIs include infection of the bladder (cystitis), kidneys

(pyelonephritis), and can lead to bacterial dissemination into the bloodstream, resulting in life-threatening urosepsis². While antibiotics remain the standard mode of treatment, their widespread use has driven the emergence of resistant strains, particularly among

¹Institute for Molecular Bioscience, The University of Queensland, Brisbane, QLD, Australia. ²Australian Infectious Diseases Research Centre, The University of Queensland, Brisbane, QLD, Australia. ³UQ Centre for Clinical Research, Faculty of Health, Medicine and Behavioural Sciences, The University of Queensland, Brisbane, Queensland, Australia. ⁴German National Reference Centre for Multidrug-resistant Gram-negative Bacteria, Department of Medical Microbiology, Ruhr-University Bochum, Universitätsstraße 150, 44801 Bochum, Germany. ⁵Pathology Queensland-Central Microbiology Laboratory, Queensland Health, Brisbane, QLD, Australia. ⁶ADVANCE-ID, Saw Swee Hock School of Public Health, National University of Singapore, Singapore, Singapore. ⁷School of Chemistry and Molecular Biosciences, The University of Queensland, Brisbane, QLD, Australia. ✉e-mail: b.forde@uq.edu.au; m.schembri@uq.edu.au

uropathogenic *E. coli* (UPEC), the causative agent of ~70% of UTIs worldwide^{2,3}. Antibiotic resistance to first-line antibiotics such as trimethoprim, co-trimoxazole, ciprofloxacin, and nitrofurantoin, as well as third-generation cephalosporins and last-line carbapenems is of particular concern⁴.

E. coli can be classified into eight broad phylogroups⁵, with phylogroups B2 and D predominantly associated with extra-intestinal infections that include UTI⁶. *E. coli* strains can be further differentiated at higher resolution using multi-locus sequence typing (MLST)⁷. This scheme has identified key UPEC sequence types (STs; also referred to as clones), including ST131, ST1193, ST95, ST73 and ST69, which are responsible for a large proportion of UTIs and bloodstream infections globally^{8–10}. These UPEC clones exhibit differences in their antibiotic resistance profiles, with ST131 and ST1193 exhibiting extensive resistance to multiple antibiotic classes⁸. The global emergence and dissemination of ST131 exemplifies the clinical impact of these clones¹¹. ST131 possesses high levels of resistance to fluoroquinolones and third-generation cephalosporins and has become the most common cause of Gram-negative extra-intestinal infections since the turn of the 21st century^{10,11}.

In contrast to these UPEC lineages, ST10 and the ST10 clonal complex (ST10cc) belong to *E. coli* phylogroup A, a group typically associated with intestinal commensal colonisation⁵. The ST10cc includes clones such as ST10, ST167, ST617 and several other STs¹². While *E. coli* strains from the ST10cc are not typically considered a major cause of extra-intestinal infections, there are reports of STs found in the ST10cc that exhibit high rates of carbapenem resistance and cause UTI and bloodstream infections^{10,13,14}.

In this study, we analysed the Enterobase¹⁵ dataset and identified an increasing number of ST167 genomes from UTI/urine samples between 2020–2024. Phylogenetic analysis of a subset of 800 ST167 genomes defined the clade structure of this high-risk clone and identified an emergent subclade (which we refer to as subclade C2) with very high carriage (89%) of genes encoding resistance to carbapenems and cephalosporins. In silico serotyping of ST167 genomic data revealed the acquisition of two group 1 capsule types within subclade C2, namely KL124^{Group1} (KL124^{G1}) and KL30^{Group1} (KL30^{G1}). We also found that ST167 strains carry significantly reduced numbers of UPEC virulence genes compared to ST131 and other UPEC strains, including an absence of genes encoding type 1 fimbriae in the majority of ST167 strains. Together, our analyses provide insight into the evolution of this emerging uropathogenic clone and identify distinct genomic signatures that support the classification of ST167 as an atypical UPEC clone associated with UTI.

Results

Identification of ST167 as a frequently sequenced clone associated with human UTI/urine

We screened Enterobase¹⁵ to detect signals of emerging *E. coli* lineages and found distinct patterns in the proportion of *E. coli* STs that cause UTIs. Data from 2024 (database accessed October 4th, 2024) revealed that ST167 represented the highest number of sequenced isolates from human UTI/urine (96 genomes), followed by ST410¹⁶ (61 genomes) and ST131¹¹ (50 genomes) (Supplementary Fig. 1; Supplementary Table 1). The two most frequently sequenced STs from all isolate sources were ST10 and ST167¹⁷, both of which belong to phylogroup A and the ST10cc¹². Strains from *E. coli* phylogroup A are typically associated with intestinal commensal colonisation⁵. ST10 showed minimal association with UTI/urine (4/371 isolates). In contrast, almost half of ST167 isolates were from UTI/urine sources (96/197 isolates), higher than the pandemic UPEC ST131 clone (50/136 isolates) (Fig. 1a). Longitudinal analysis of UTI/urine isolates in Enterobase from the five most deposited STs revealed a recent increase of ST167. ST167 increased ~17-fold in the Enterobase dataset, from 1.2% of total UTI/urine isolates in 2020 to 21.2% in 2024 (Fig. 1a; Supplementary Table 2). The increase in

ST167 depositions to Enterobase, coupled with the high number of sequenced isolates from UTI/urine sources, suggests this ST represents an emergent phylogroup A uropathogenic clone.

Genomic phylogeny of ST167 reveals an emergent subclade with increased antibiotic resistance

To better understand the evolution and recent expansion of ST167, we analysed 2,066 ST167 genomes from Enterobase¹⁵. Initial phylogenetic analysis with Mashtree¹⁸ revealed five putative clades in the ST167 phylogeny (Supplementary Fig. 2). Subsequently, we constructed a higher resolution maximum likelihood phylogenetic tree using a randomly selected subset of 800 ST167 genomes, based on 109,894 recombination-free core-genome polymorphic sites (Fig. 1b). Bayesian analysis of population structure (fastbaps)¹⁹ identified 17 distinct clusters at BAPS level 1 and level 2 (Supplementary Fig. 3). The three largest clusters were defined as clade A (70 genomes), B (314 genomes) and C (329 genomes), with the 14 remaining clusters ranging in sizes from 1–23 genomes (Supplementary Fig. 3; Supplementary Data 1). Clade C was further resolved into six subclades; subclade C1 and subclade C2 were the largest with 38 genomes and 262 genomes, respectively (Supplementary Fig. 3; Supplementary Data 1). Notably, 84% of subclade C2 strains were isolated after 2020 (Fig. 1c), while only 34% of clade B strains were isolated in the same period, suggesting the recent emergence of subclade C2 within the ST167 population. Further analysis of the isolation source revealed that subclade C2 had a significantly higher association with UTIs and urine samples (109/262 isolates [41.6%]) compared to clade A (10/70 isolates [14.3%]; $P = 6.8 \times 10^{-5}$), clade B (60/314 isolates [19.1%]; $P = 3.1 \times 10^{-8}$) and subclade C1 (7/38 isolates [18.4%]; $P = 4.16 \times 10^{-2}$; Fisher's exact test with Bonferroni-adjusted p -values) (Fig. 1d). This suggests that isolates from subclade C2 may exhibit increased pathogenic potential, providing valuable insights into the evolving epidemiology and potential clinical implications of the ST167 lineage.

The carriage of genes conferring resistance to antibiotics, including cephalosporins and last line carbapenems, is of clinical importance for treatment options. Among genes that conferred resistance to these classes of antibiotics, *bla*_{CTX-M-15} (1186/2066) and *bla*_{NDM-5} (1330/2066) were most common in ST167 genomes (Supplementary Data 2). Seventy four percent of genomes in clade B and 89% in subclade C2 possessed genes conferring resistance to both cephalosporins and carbapenems (Fig. 1e). To validate the in silico resistance profiles, we sourced strains from clade B ($n = 6$) and subclade C2 ($n = 5$) and conducted susceptibility testing for 24 antibiotics across 9 different classes. Most isolates from clade B and C2 were non-susceptible to cephalosporins and carbapenems, while there were differences in susceptibility for the isolates examined to amikacin, gentamicin and minocycline (Supplementary Data 3). All isolates examined were susceptible to nitrofurantoin (Supplementary Data 3).

ST167 lack many UPEC-associated virulence factors, including type 1 fimbriae

To investigate the virulence potential of ST167, we conducted a comprehensive analysis of UPEC virulence genes across 800 genomes from our phylogenetic dataset. We curated a list of 74 non-redundant marker genes for established UPEC virulence factors, encompassing genes involved in adhesion, iron acquisition, immune evasion, motility and toxin production (Supplementary Data 4). We first calculated a virulence factor score for ST167 and phylogroup B2 and D STs strongly associated with UTI and demonstrated that ST167 has a reduced number of virulence factors compared to these UPEC clones (Supplementary Fig. 4). Comparative genomic analyses between ST167 and ST131 revealed that ST167 had significantly fewer UPEC virulence factors corresponding to chaperone-usher fimbriae, autotransporters, iron uptake, toxins and immune evasion mechanisms (Supplementary Fig. 5). The distribution of virulence factors within the phylogenetic

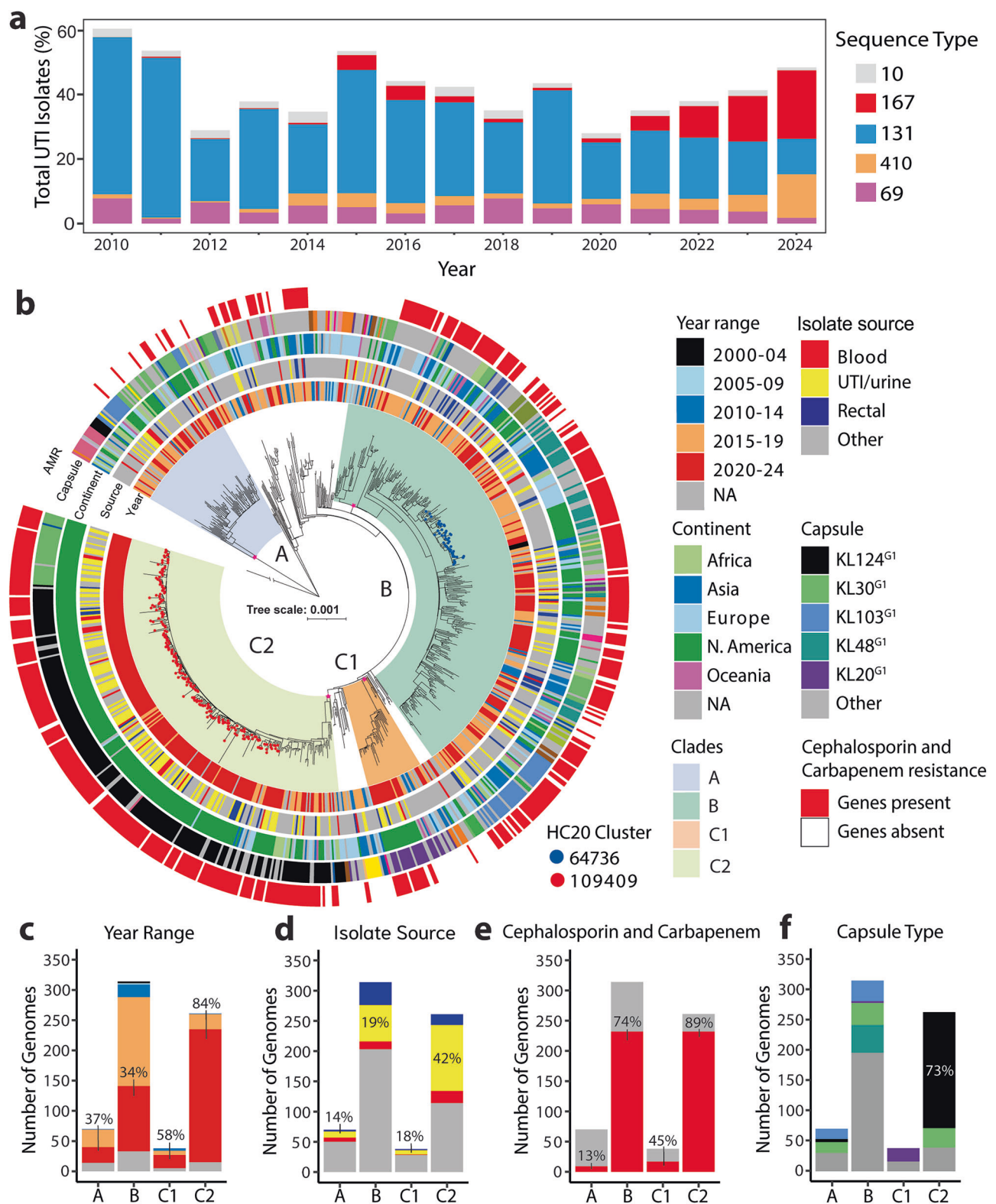


Fig. 1 | Genomic analysis of ST167. a Bar graph displaying the human UTI/urine isolate proportion (%) from 2010 to 2024 for the five most sequenced *E. coli* STs deposited to Enterobase in 2024 (source data in Supplementary Table 2). **b** A Maximum likelihood tree of 800 ST167 genomes. The tree was constructed from 109,894 recombination-free core-genome polymorphic sites by IQ-TREE (GTR + G model and 1000 ultrafast bootstraps; with CP083869.1 used as reference and ED1a as outgroup). Tree scale is representative of nucleotide substitutions per site. ST167 clades A, B and subclades C1 and C2 are shaded and were determined using fastbaps. The HC20 clusters are coloured on tip points of the branches. The pink stars

on the branches of the major clade are indicative of >90% bootstrap support. The capsule type of the ST167 strains was determined using Kaptive 2.0. Antibiotic resistance genes carried by the ST167 strains was determined using AMRfinderplus, with red on the outermost ring indicating carriage of both carbapenem and cephalosporin resistance genes. **c–f** Enterobase metadata was compiled to display bar charts for (c) collection year, (d) isolate source, 10/70 isolates were from UTI/urine sources for clade A, 60/314 for clade B, 7/38 for subclade C1 and 109/262 for C2, (e) carbapenem and cephalosporin resistance, and (f) capsule type.

structure of ST167 also revealed clade-specific patterns; for example, the presence of a specific *agn43* (antigen 43) allele in subclade C2 and absence of the yersiniabactin receptor gene *fyuA* in subclade C2 (Fig. 2a). Analysis of chaperone-usher fimbriae revealed the presence of genes encoding Ecp/Mat fimbriae (*ecpC*), Yad fimbriae (*yadV*) and Yfc fimbriae (*yfcV*) in the majority of strains, and Yeh fimbriae (*yehC*) in all ST167 except clade A strains. Strikingly, almost all ST167 strains except those from clade A lacked type 1 fimbriae (*fimD*). Further investigation into the absence of type 1 fimbriae genes revealed a complete replacement of the entire *fim* locus by the *ISIA* insertion element (Fig. 2b). Type 1 fimbriae are a critical UPEC virulence factor required for bladder colonisation^{20,21}. Type 1 fimbriae-mediated binding via the tip-located FimH adhesin facilitates UPEC adhesion to α -D-mannosylated glycoproteins such as uroplakins on the bladder epithelium²², facilitating invasion of bladder superficial epithelial cells²³ and the formation of intracellular bacterial communities (IBCs)²⁴. To assess whether ST167 subclade C2 could colonise the bladder despite the lack of type 1 fimbriae, we employed a murine UTI model. ST167 subclade C2 strains MS25298 and MS25303 were recovered at variable loads in the urine, with bacteria detected in 11/19 and 5/19 mice, respectively. In contrast, the reference UPEC strain CFT073 belonging to ST73 was recovered in significantly higher bacterial loads in the urine (Fig. 2c). Bladder colonisation of MS25298 and MS25303 was also significantly reduced compared to CFT073; MS25298 and MS25303 were detected in 1/19 and 7/19 mice at $\sim 10^4$ CFU/g tissue, respectively, compared to CFT073 which effectively colonised the bladder of all mice at $\sim 10^7$ CFU/g tissue (Fig. 2d). Neither the ST167 strains nor CFT073 effectively colonised the kidneys (Fig. 2e), consistent with previous data from our lab using CFT073 for infection of C57BL/6 mice^{25,26}. Taken together, our discovery that type 1 fimbriae are absent in most ST167 leads us to propose ST167 as an atypical UPEC clone associated with UTI.

Core genome MLST reveals two closely related ST167 clusters

Evidence for the recent expansion of ST167 subclade C2 prompted a closer investigation into the genomic similarity of the ST167 genomes. To assess this, we extracted the hierarchical clusters of core gene multi-locus sequence typing (HierCC)²⁷ data at the HC20 level (reflecting genomes clustered with <20 allelic differences in 2513 core genes) from Enterobase¹⁵. The ST167 population revealed two HC20 clusters that contained >100 genomes each (Fig. 1a). Clade B contained the HC20 cluster 64736, which was primarily isolated from companion animals (75/176 isolates). Subclade C2 contained HC20 cluster 109409 which consisted almost exclusively of isolates from humans (419/433; Fig. 3a, b). Strains in the two HC20 clusters possessed high rates of carriage of *bla*_{NDM-5} (carbapenem resistance), *bla*_{CTX-M-15} and *bla*_{OXA-1} (cephalosporin resistance) (Fig. 3c). In contrast, carriage of other antibiotic resistance genes differed between the HC20 clusters (Fig. 3c). HC20 cluster 64736 showed high proportions of *bla*_{TEM-1} (resistance to penicillins), *dfrA17* (trimethoprim resistance), *aac*(3)-IId (gentamicin resistance), as well as *aadA5* and *aadA22* (streptomycin resistance). Cluster 109409 contained a higher proportion of *sul2* (sulfonamide resistance), *aph*(3')-Ia (kanamycin resistance), *aph*(6)-Id and *aph*(3'')-Ib (streptomycin resistance) genes.

Serotype diversity within ST167 is determined by changes in capsule and flagella antigens

In silico serotyping analysis revealed differences in the capsule (K)- and flagella (H)-antigen among the HC20 clusters, while the O89b O-antigen was shared by strains in both HC20 clusters and the majority of other ST167 strains (Supplementary Fig. 6a). The O89b O-antigen is associated with group 1 capsule in *E. coli*^{28,29}. Given the high levels of similarity between the group 1 capsule of *E. coli* and *Klebsiella pneumoniae*³⁰, we employed Kaptive 2.0³¹, a *K. pneumoniae* capsule typing tool. Capsule typing identified 30 group 1 capsule types in

ST167 (Fig. 1f; Supplementary Data 2). The group 1 capsule is located between *galF* and *ugd* on the chromosome; the dominant HC20 cluster 109409 primarily contained the KL124^{G1} capsule type (Fig. 3d) and a smaller proportion of the KL30^{G1} capsule type (Supplementary Fig. 6b). HC20 cluster 64736 contained the KL48^{Group1} (KL48^{G1}) capsule type (Supplementary Fig. 6b) that has also been identified in carbapenem resistant ST167³². H-antigen typing revealed HC20 cluster 64736 contained the H9 H-antigen, while 109409 harboured the H10 H-antigen (Supplementary Fig. 6a). These data demonstrate that subclade C2 is represented by a HC20 cluster with two distinct serotypes, KL124^{G1}:O89b:H10 and KL30^{G1}:O89b:H10.

Commonly, serotype diversity is determined by recombination events at the capsule, O-antigen and flagella loci. Acquisition of new genomic features through recombination has previously been linked to the expansion of other UPEC clones^{33,34}. To investigate if ST167 subclade C2 acquired new genomic features via recombination we assessed recombination events using Gubbins³⁵. This analysis identified 11 recombination regions >10 kb associated with subclade C2. The largest recombination region was ~387 kb in length and contained the flagella and capsule genes (region E; Supplementary Fig. 7). Other recombination regions contained genes associated with phage, genomic islands and core genes, including regions B (150 kb) and C (86 kb) that were restricted to ST167 subclade C2 (Supplementary Data 5).

Bayesian analysis of ST167 subclade C2 identifies a rapidly expanding cluster comprising two distinct capsule types

The geographical distribution of subclade C2 shows that 87.6% of the strains were isolated in North America, associated with the expansion of the HC20 cluster 109409 which composed ~77% of subclade C2 isolates (Fig. 4a). To assess the population dynamics of subclade C2 we extracted all subclade C2 genomes from Enterobase ($n = 539$) and assessed the temporal signal with TempEST³⁶. There was evidence for a modest clock-like evolution (correlation coefficient = 0.3252; $R^2 = 0.1057$). The identification of the clock-like structure of subclade C2 was constrained by limited metadata detailing collection date, with only year of isolation consistently available across isolates. Despite these constraints, the positive temporal signal allowed us to employ BEAST³⁷ to reconstruct the evolutionary timeline of subclade C2. We determined the best-fitting model to be the uncorrelated relaxed exponential clock with the Bayesian skyline population model based upon the mean tree likelihood (Supplementary Data 6). The time to most recent common ancestor (TMRCA) of the HC20 cluster was 2015.8 (95% highest posterior density (HPD): 2014.9–2016.5) (Fig. 4a). The expansion of the HC20 cluster 109409 correlated with the Bayesian skyline plot that showed a clonal expansion around 2016 (Fig. 4b). Taken together, these data suggest HC20 cluster 109409 is an expanding clone with two distinct serotypes, primarily sequenced in the United States and associated with UTI.

Discussion

Here we report on ST167 as an emergent carbapenem-resistant atypical UPEC clone associated with UTI. By generating a comprehensive ST167 phylogeny, we identified the acquisition of group 1 capsule genes and the loss of type 1 fimbriae genes as key features of most ST167 strains.

This study had several limitations. The genomic data was sourced from Enterobase, a publicly available database that primarily obtains its data from the NCBI Short Read Archive (SRA). Although Enterobase offers valuable insights into genomic epidemiology, its use of publicly submitted sequencing data reflects the sequencing activities and priorities of contributing researchers rather than a structured surveillance system. Consequently, the data does not provide a comprehensive representation of all bacterial populations or geographic regions, which may affect the generalisability of the study's findings.

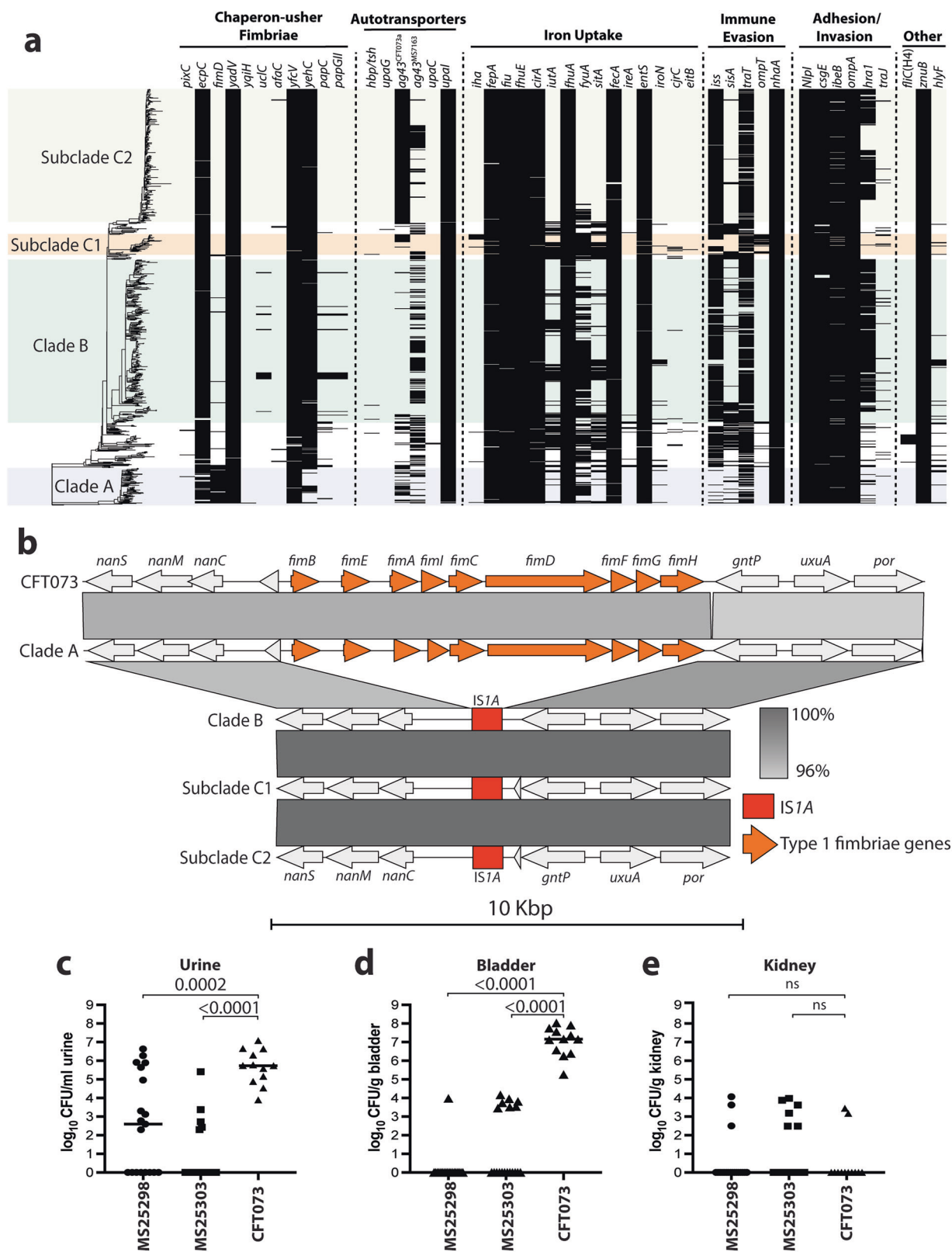


Fig. 2 | Virulence analysis of ST167. a Virulence factor heatmap of ST167. On the left is the phylogenetic tree of ST167 depicting the 800 genomes shown in Fig. 1a. The heatmap shows the presence and absence of the virulence factor marker genes; black indicates the virulence factor marker gene is present. The defined clades of the ST167 phylogeny are shaded. **b** EasyFig comparison of the type 1 fimbriae locus in CFT073 and representative complete genomes from clades A, B, C1 and C2. **c–e** Bacterial loads recovered from the (c) urine, (d) bladder and (e) kidneys of infected C57BL/6 mice expressed as median bacterial load per gram of tissue or ml

of urine. Data are pooled from 2 independent experiments ($n = 19$ for strains MS25298 and MS25303; $n = 12$ for CFT073). Each symbol represents the bacterial load from an individual mouse at 24 h post infection. MS25298 and MS25303 are subclade C2 genomes from HC20 cluster 109409 and contain the KL124 capsule. Statistical significance was determined by an ordinary one-way ANOVA with the Dunnett correction applied; MS25298 and MS25303, respectively, were compared to CFT073.

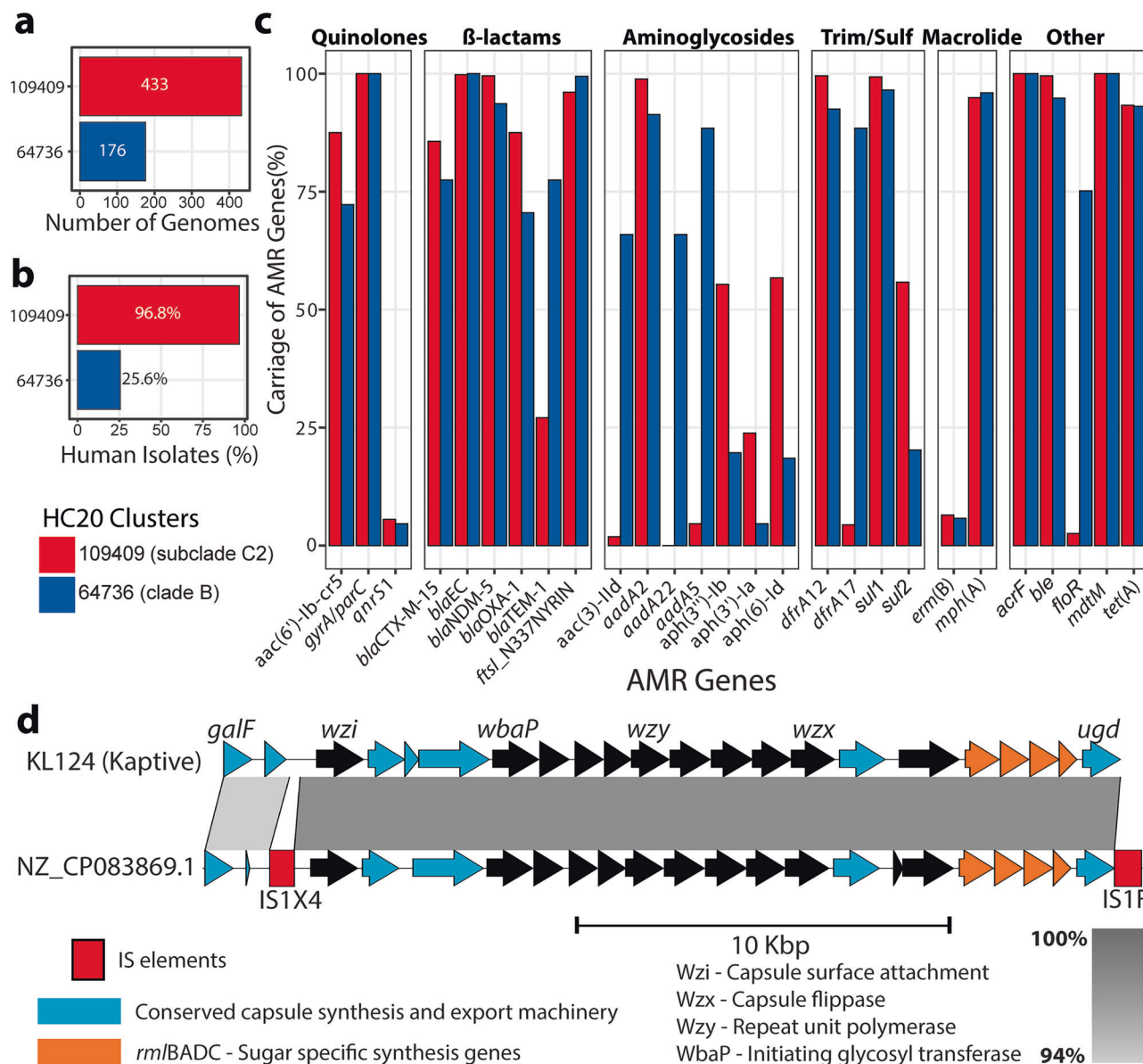


Fig. 3 | Antibiotic resistance and capsule loci analysis of ST167 HC20 clusters. a Bar graph depicting the number of genomes within each HC20 cluster. The colour is representative of HC20 cluster, red is 109409 and blue is 64736. **b** Bar graph showing the percentage of genomes isolated from human sources from each HC20 cluster. **c** Bar graph demonstrating the percentage of genomes from each HC20 cluster that carry the respective antibiotic resistance genes displayed on the x-axis (source data in Supplementary Data 2). The category of Trim/Sulf is representative

of trimethoprim and sulfonamide antibiotic classes. The annotation of *gyrA/parC* denotes the chromosomal mutations of *gyrA*_D87N, *gyrA*_S83L, *parC*_S80I and *parE*_S458A. **d** Sequence comparison of KL124⁶¹ Kaptive and NZ_CP083869.1 as a representative strain, visualised with Easyfig; the level of nucleotide sequence conservation is indicated by shading based on the legend. Capsule locus genes were annotated using Kaptive reference gene locus tags. IS elements were identified with ISfinder.

We note that the majority of ST167 North American UTI/urine isolates (88.3%) were collected as part of the Centers for Disease Control and Prevention (CDC) National Healthcare Safety Network (NHSN). The NHSN is a large programme that conducts surveillance for healthcare events across diverse facilities in the United States, including the collection of data and biological samples associated with healthcare-associated infections (HAIs) and antimicrobial resistance³⁸. Leveraging the NHSN extensive surveillance data, we inferred an increase in the proportion of genome sequenced UTI-associated *E. coli* ST167 in North America. However, the high number of genomes contributed to Enterobase through this programme means there are sampling biases in the data, for example the overrepresentation of human pathogens that: (i) are sequenced from high resource settings, (ii) cause HAIs, and

(iii) possess resistance to critical antibiotics such as carbapenems. Thus, while our findings identify ST167 subclade C2 as a frequently sequenced UPEC clone that causes UTI, we do not have sufficient data to compare infection incidence to other global clones such as ST131. Despite this bias, our analysis supports a distribution of ST167 beyond North America, with detection of subclade C2 isolates from geographic regions spanning Europe, Africa and Asia (Supplementary Fig. 8; Supplementary Data 2).

The geographical pattern we observed reflecting an increasing number of ST167 genomes with carbapenem resistance isolated in North America is also supported by the NCBI pathogen detection tool. Here, SNP cluster PDS000063179.375 ($n = 528$ isolates), which encompasses HC20 cluster 109409 of subclade C2, shows while 86.4%

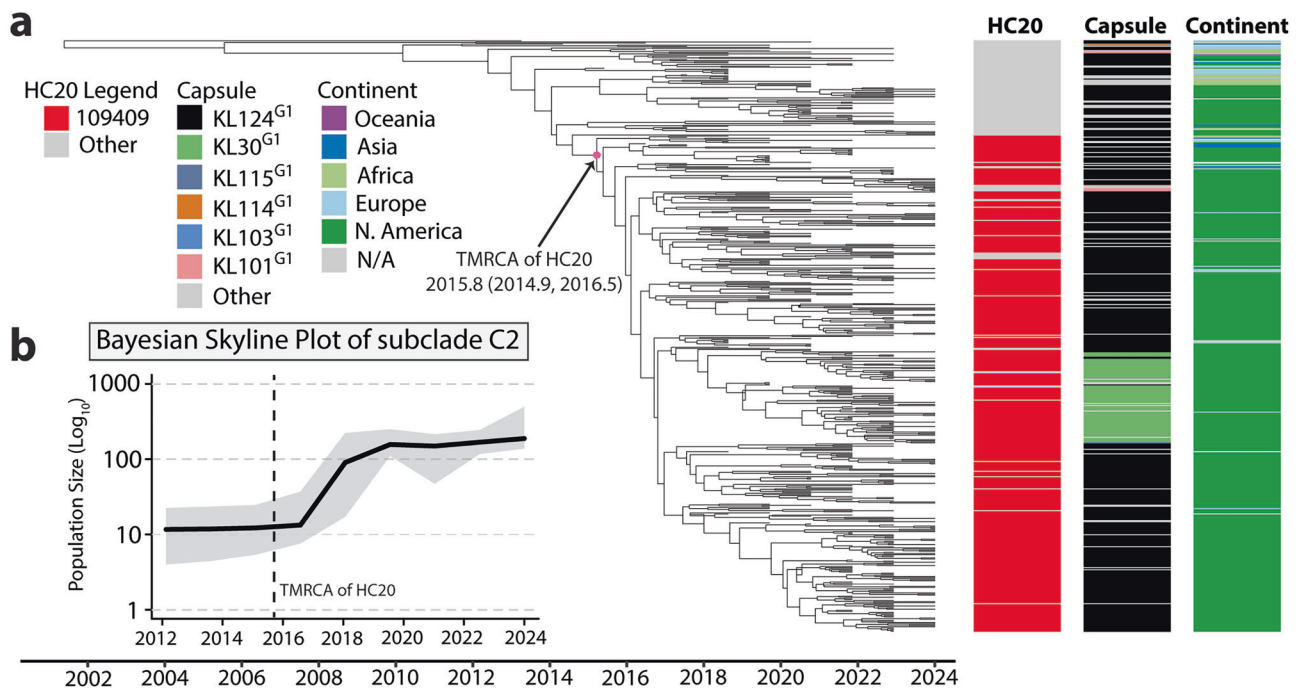


Fig. 4 | Bayesian analysis of ST167 subclone C2. a Time scaled tree of ST167 subclone C2 consisting of 539 genomes. The pink polygon depicts the node correlating to the TMRCA of HC20 cluster 109409. Metadata identifying HC20 cluster, capsule type and continent of isolation are displayed on the right as

columns. **b** Bayesian skyline plot of subclone C2 genomes with the dotted line corresponding to the TMRCA of the HC20 cluster 109409. The black curve represents the effective population size, with the grey shading representing the 95% confidence interval.

of isolates originate from this region, isolates were also detected from 18 countries in other geographic regions (Supplementary Fig. 9). However, detailed assessment of infection incidence of subclone C2 in these other regions is limited by the availability and scope of comparable surveillance datasets. To further address potential database-driven biases, we conducted a meta-analysis of published studies from 2020 to 2025 that reported MLST data on carbapenem-resistant *E. coli* isolates, with a threshold of at least 50 isolates required for inclusion. The 33 studies included in this analysis varied widely in sample size and focus, with uneven geographical representation across regions, which limits the generalisability of the findings. Despite these constraints, the meta-analysis revealed that ST131 was the most frequently detected carbapenem-resistant clone, with ST167 detected at comparable levels to ST410, ST405 and ST38 (Supplementary Data 7).

In conclusion, our genomic investigation provides insights into the emergence of ST167 as a multidrug resistant UPEC clone associated with UTI, an observation that builds on other reports linking ST167 with increasing carbapenem resistance^{13,39,40}. A striking finding from our genomic investigation of ST167 was the complete loss of genes encoding type 1 fimbriae in almost all ST167 strains from clade B, C1 and C2. Given the critical role of type 1 fimbriae for UPEC bladder colonisation in mouse models that replicate UTI^{20,21}, the inability to produce these fimbriae suggests ST167 disease pathogenesis is different to previously characterised phylogroup B2 and D UPEC strains and associated with alternative mechanisms of colonisation and/or virulence factors. The group 1 capsule identified in ST167, likely originating from *Klebsiella*, may be one such virulence factor. Our discovery also raises important considerations concerning ongoing UTI programmes aimed at the development of new anti-virulence drugs to block binding of the type 1 fimbriae FimH tip adhesin to the bladder uroepithelium⁴¹, or vaccination with FimH⁴², as such approaches would be ineffective at treating or preventing UTI caused by ST167. Together, these considerations form the basis of our criteria for the classification of ST167 as an atypical UPEC clone.

Methods

Ethics approval

Mouse infection experiments were approved by the University of Queensland Animal Ethics Committee (AEC approval number SCMB/259/19).

Enterobase data

The *Escherichia/Shigella* dataset from Enterobase¹⁵ was downloaded on the 4th of October, 2024. *Shigella* STs were identified and filtered out using a separate *Shigella*-specific dataset from Enterobase¹⁵. To determine STs associated with UTIs, the *E. coli* dataset was first filtered for human isolates (search terms: 'Human|Homo sapiens') and subsequently for urinary sources (search terms: 'urine|UTI'). An ST with at least one isolate associated with human UTI identified using this method was considered a human UTI-associated ST. The human UTI-associated genomes were then selected from the entire dataset to identify the proportion of human UTI-associated STs. Data detailing the continent of origin for genomes and number of contributing sequencing projects per year is provided in Supplementary Data 8.

Phylogenetic analysis

To construct a phylogeny for ST167, 2,104 draft genome assemblies were downloaded from Enterobase¹⁵. Quality control was performed with QUAST⁴³ v5; genomes with >500 contigs and >20 undetermined bases per 100 Kb were removed, yielding a dataset of 2,066 genomes. A phylogenetic tree was constructed by MashTree¹⁸ v1.4.6 to understand the overall topology of ST167 (Supplementary Fig. 2).

To construct a maximum likelihood phylogenetic tree of ST167, 800 genomes were randomly selected in R. Prokka⁴⁴ v1.13 was then used for genome annotation. Core genome SNPs were called with snippy and snippy-core v4.6.0 (<https://github.com/tseemann/snippy>) generated a pseudo-genome alignment with the SNPs reinserted into the reference strain CP083869.1 genome (accession number CP083869.1); ED1a (accession number CU928162) served as an

outgroup for phylogenetic rooting. Regions of recombination were detected using Gubbins³⁵ v3.2.1 under default settings. Predicted recombinant regions were removed and the output was used to generate a phylogenetic tree with IQ-TREE⁴⁵ v2.1.2 using the general time-reversal GAMMA model (GTR + G) and 1000x ultrafast bootstraps⁴⁶, all other settings were default. Phylogenetic clade structure was determined using Bayesian analysis of population structure (fastbaps)⁴⁹ implemented in the R package fastbaps v1.0.8. Clade A was determined at the first level, clade B and C at the second level and subclades of clade C at the third level (Supplementary Fig. 3; Supplementary Data 1). The phylogenetic tree was visualised in iTOL⁴⁷.

Accessory genome analysis

To determine the serotype of the ST167 genomes, we first performed capsule typing with Kaptive 2.0 software using the *Klebsiella* k locus primary reference database³¹ (downloaded 27/09/24). Genomes with confidence of 'None' and 'Low' from the Kaptive output were considered untypeable. To confirm the similarity of group 1 capsule in *E. coli* and *Klebsiella*, we performed pairwise comparison between completely sequenced genomes from RefSeq and the Kaptive database reference loci using Easyfig⁴⁸ v2.2.2. The EcOH database (downloaded 10/08/2023) from the ABRicate software v1.0.1 (<https://github.com/tseemann/abricate>) was used to determine O-antigen and H-antigen type⁴⁹. Analysis of antibiotic resistance gene carriage in ST167 genomes was performed with AMRfinderplus⁵⁰ v3.11.20. PHASTEST⁵¹ was used to identify phage regions within the reference strain CP083869.1. Phandango⁵² was used to visualise recombination regions. ISfinder⁵³ was used to identify IS elements.

Virulence factor analysis

To identify UPEC-associated virulence factors we curated a list of established UPEC virulence factor marker genes from the literature (Supplementary Data 4). This virulence factor list contained 74 non-redundant marker genes. Comparative analysis with the ST131 was performed using a previously published ST131 dataset containing 3,993 genomes⁵⁴.

Antibiotic susceptibility testing

Antibiotic susceptibility testing of ST167 isolates was performed using broth microdilution (Sensititre; Thermo Fisher Scientific) and results were interpreted according to European Committee on Antimicrobial Susceptibility Testing (EUCAST) clinical breakpoints criteria V13.0⁵⁵. For antibiotics that do not have a defined breakpoint according to EUCAST guidelines (e.g. minocycline) we deferred to Clinical and Laboratory Standards Institute (CLSI) guidelines⁵⁶. Nitrofurantoin susceptibility was determined using the breakpoint for uncomplicated UTI⁵⁵.

Mouse infection experiments and antibiogram testing

All experiments were performed using the following housing conditions: (light:dark cycle 12:12 h, room temperature $21 \pm 1^\circ\text{C}$, humidity $50 \pm 10\%$). The mouse UTI model was performed essentially as previously described^{57,58}. Strains were cultured overnight in LB broth with shaking for inoculum preparation. Female C57BL/6 mice (8–10 weeks) were infected via transurethral inoculation at 5×10^8 CFU/ml. Mice were euthanised by cervical dislocation at 24 h post infection. Bacterial loads corresponding to each strain in the urine, bladder and kidney of infected mice were enumerated by plating onto LB agar. All experiments were conducted in duplicate using groups of 9–10 mice (MS25298 and MS25303) or 6 mice (CFT073). Strains MS25298 and MS25303 were tested in a total of 19 mice, respectively; CFT073 was tested in a total of 12 mice.

Bayesian temporal analysis

The HC100 cluster 71026 from EnteroBase²⁷ was identified to encompass the entirety of subclade C2 genomes and was extracted. After

filtering with QUAST⁴³ there was a collection of 539 ST167 subclade C2 genomes with metadata for collection year. Preliminary estimation of the temporal signal was determined by TempEST³⁶ by performing a regression analysis of root-to-tip genetic distance extracted from a recombination-free maximum-likelihood phylogenetic tree. The IQtree⁴⁵ model selection function was used to identify the nucleotide substitution model for BEAST³⁷ and was determined to be a GTR + G model. The Gamma Site Model Category Count was set to four and the GTR substitution model rates determined from IQtree⁴⁵ were included (rate AC = 1.0715, AG = 2.9806, AT = 1.8727, CG = 0.1453, CT = 3.2344, and GT = 1.00). The initial clock rate was set to 5.3222×10^{-4} as estimated by the root-to-tip regression analysis in TempEST³⁶. To identify the optimal combination for molecular clock and population growth models, we evaluated a number of combinations with BEAST v2.6.7. These included the strict, log-normal and relaxed exponential clock models, paired with constant, exponential and bayesian skyline population models. Markov chain Monte Carlo (MCMC) generations for each analysis were conducted in triplicate for 100 million steps, sampling every 1000 to ensure convergence. Log-combiner was used to collate the log files with a 10% burn-in and Tracer v1.7.2 was used to assess the BEAST runs. The best fitting model was determined by the mean tree likelihood and was found to be the uncorrelated relaxed exponential clock with the Bayesian skyline population model (Supplementary Data 6). TreeAnnotator was used to generate the maximum clade credibility (MCC) tree with a 0.5 posterior probability limit.

Meta-analysis of CREC isolates

To identify studies reporting carbapenem-resistant *E. coli* we searched PubMed with the search term 'carbapenem* resistant *E. coli* sequence type*'. The inclusion of studies required publication between 1st of January 2020 to the 27th of March 2025 and MLST analysis on a dataset of ≥ 50 isolates. These criteria resulted in identification of 33 publications that encompassed 6,877 carbapenem *E. coli* isolates (Supplementary Data 7).

Statistics and reproducibility

Statistical analyses of bacterial loads in the urine, bladder and kidney of infected mice were performed in GraphPad Prism (version 10.2.2). Other statistical analyses to determine significance for proportion of UTI isolates and carriage of virulence factors were performed in R. All information on statistical tests and sample sizes are described in the appropriate figure legends and methods sections. No statistical method was used to determine sample size. No data was excluded from the analyses. The investigators were not blinded to allocation during experiments and assessment.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

All data generated in this work is presented in the manuscript. Source data are provided with this paper.

Code availability

The code used in this manuscript to generate figures presented have been made available in the following repository - <https://doi.org/10.5281/zenodo.16756974>.

References

1. Yang, X. et al. Disease burden and long-term trends of urinary tract infections: a worldwide report. *Front. Public Health* **10**, 888205 (2022).

2. Flores-Mireles, A. L., Walker, J. N., Caparon, M. & Hultgren, S. J. Urinary tract infections: epidemiology, mechanisms of infection and treatment options. *Nat. Rev. Microbiol.* **13**, 269–284 (2015).
3. Foxman, B. Urinary tract infection syndromes: occurrence, recurrence, bacteriology, risk factors, and disease burden. *Infect. Dis. Clin. North Am.* **28**, 1–13 (2014).
4. Walker, M. M., Roberts, J. A., Rogers, B. A., Harris, P. N. A. & Sime, F. B. Current and emerging treatment options for multidrug resistant *Escherichia coli* urosepsis: a review. *Antibiotics (Basel)*. **11**, 1821 (2022).
5. Denamur, E., Clermont, O., Bonacorsi, S. & Gordon, D. The population genetics of pathogenic *Escherichia coli*. *Nat. Rev. Microbiol.* **19**, 37–54 (2021).
6. Sannes, M. R., Kuskowski, M. A., Owens, K., Gajewski, A. & Johnson, J. R. Virulence factor profiles and phylogenetic background of *Escherichia coli* isolates from veterans with bacteremia and uninfected control subjects. *J. Infect. Dis.* **190**, 2121–2128 (2004).
7. Tartof, S. Y., Solberg, O. D., Manges, A. R. & Riley, L. W. Analysis of a uropathogenic *Escherichia coli* clonal group by multilocus sequence typing. *J. Clin. Microbiol.* **43**, 5860–5864 (2005).
8. Fibke, C. D. et al. Genomic epidemiology of major extraintestinal pathogenic *Escherichia coli* lineages causing urinary tract infections in young women across Canada. *Open Forum Infect. Dis.* **6**, ofz431 (2019).
9. Lipworth, S. et al. Ten-year longitudinal molecular epidemiology study of *Escherichia coli* and *Klebsiella* species bloodstream infections in Oxfordshire, UK. *Genome Med.* **13**, 144 (2021).
10. Manges, A. R. et al. Global extraintestinal pathogenic *Escherichia coli* (ExPEC) lineages. *Clin. Microbiol. Rev.* **32**, e00135–18 (2019).
11. Petty, N. K. et al. Global dissemination of a multidrug resistant *Escherichia coli* clone. *Proc. Natl. Acad. Sci. USA* **111**, 5694–5699 (2014).
12. Kocsis, B., Gulyás, D. & Szabó, D. Emergence and dissemination of extraintestinal pathogenic high-risk international clones of *Escherichia coli*. *Life (Basel)*. **12**, 2077 (2022).
13. Huang, J. et al. Carbapenem-resistant *Escherichia coli* exhibit diverse spatiotemporal epidemiological characteristics across the globe. *Commun. Biol.* **7**, 51 (2024).
14. Kallonen, T. et al. Systematic longitudinal survey of invasive *Escherichia coli* in England demonstrates a stable population structure only transiently disturbed by the emergence of ST131. *Genome Res.* **27**, 1437–1449 (2017).
15. Zhou, Z., Alikhan, N. F., Mohamed, K., Fan, Y. & Achtman, M. The Enterobase user's guide, with case studies on *Salmonella* transmissions, *Yersinia pestis* phylogeny, and *Escherichia coli* core genomic diversity. *Genome Res.* **30**, 138–152 (2020).
16. Ba, X. et al. Global emergence of a hypervirulent carbapenem-resistant *Escherichia coli* ST410 clone. *Nat. Commun.* **15**, 494 (2024).
17. Garcia-Fernandez, A. et al. Novel insights and features of the NDM-5-producing *Escherichia coli* sequence type 167 high-risk clone. *mSphere* **5**, e00269–20 (2020).
18. Katz, L. S. et al. MashTree: a rapid comparison of whole genome sequence files. *J. Open Source Softw.* **4**, 10.21105/joss.01762 (2019).
19. Tonkin-Hill, G., Lees, J. A., Bentley, S. D., Frost, S. D. W. & Corander, J. Fast hierarchical Bayesian analysis of population structure. *Nucleic Acids Res.* **47**, 5539–5549 (2019).
20. Connell, I. et al. Type 1 fimbrial expression enhances *Escherichia coli* virulence for the urinary tract. *Proc. Natl. Acad. Sci. USA* **93**, 9827–9832 (1996).
21. Mulvey, M. A. et al. Induction and evasion of host defenses by type 1-piliated uropathogenic *Escherichia coli*. *Science* **282**, 1494–1497 (1998).
22. Wu, X. R., Sun, T. T. & Medina, J. J. In vitro binding of type 1-fimbriated *Escherichia coli* to uroplakins Ia and Ib: relation to urinary tract infections. *Proc. Natl. Acad. Sci. USA* **93**, 9630–9635 (1996).
23. Martinez, J. J., Mulvey, M. A., Schilling, J. D., Pinkner, J. S. & Hultgren, S. J. Type 1 pilus-mediated bacterial invasion of bladder epithelial cells. *EMBO J.* **19**, 2803–2812 (2000).
24. Anderson, G. G. et al. Intracellular bacterial biofilm-like pods in urinary tract infections. *Science* **301**, 105–107 (2003).
25. Ulett, G. C. et al. Functional analysis of antigen 43 in uropathogenic *Escherichia coli* reveals a role in long-term persistence in the urinary tract. *Infect. Immun.* **75**, 3233–3244 (2007).
26. Allsopp, L. P. et al. Molecular characterization of UpaB and UpaC, two new autotransporter proteins of uropathogenic *Escherichia coli* CFT073. *Infect. Immun.* **80**, 321–332 (2012).
27. Zhou, Z., Charlesworth, J. & Achtman, M. HierCC: a multi-level clustering scheme for population assignments based on core genome MLST. *Bioinformatics* **37**, 3645–3646 (2021).
28. Harris, S. et al. Variant O89 O-antigen of *E. coli* is associated with group 1 capsule loci and multidrug resistance. *Front Microbiol.* **9**, 2026 (2018).
29. Zeng, X. et al. Comparative genome analysis of an extensively drug-resistant isolate of avian sequence type 167 *Escherichia coli* strain sanji with novel in silico serotype O89b:H9. *mSystems* **4**, e00242–18 (2019).
30. Rahn, A., Drummelsmith, J. & Whitfield, C. Conserved organization in the *cps* gene clusters for expression of *Escherichia coli* group 1 K antigens: relationship to the colanic acid biosynthesis locus and the *cps* genes from *Klebsiella pneumoniae*. *J. Bacteriol.* **181**, 2307–2313 (1999).
31. Lam, M. M. C., Wick, R. R., Judd, L. M., Holt, K. E. & Wyres, K. L. Kaptive 2.0: updated capsule and lipopolysaccharide locus typing for the *Klebsiella pneumoniae* species complex. *Microbial. Genomics* **8**, 000800 (2022).
32. Wang, M. et al. KpnK48 clone driving hypervirulent carbapenem-resistant *Escherichia coli* epidemics: Insights into its evolutionary trajectory similar to *Klebsiella pneumoniae*. *Drug Resist. Updat.* **81**, 101243 (2025).
33. Tchesnokova, V. et al. Pandemic fluoroquinolone resistant *Escherichia coli* clone ST1193 emerged via simultaneous homologous recombinations in 11 gene loci. *Proc. Natl. Acad. Sci. USA* **116**, 14740–14748 (2019).
34. Zakour, N. L. B. et al. Sequential acquisition of virulence and fluoroquinolone resistance has shaped the evolution of *Escherichia coli* ST131. *mBio* **7**, 00347–00316 (2016).
35. Croucher, N. J. et al. Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. *Nucleic Acids Res.* **43**, e15–e15 (2014).
36. Rambaut, A., Lam, T. T., Max Carvalho, L. & Pybus, O. G. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evol.* **2**, vew007 (2016).
37. Bouckaert, R. et al. BEAST 2: A software platform for bayesian evolutionary analysis. *PLOS Comput. Biol.* **10**, e1003537 (2014).
38. Centre for Disease Control and Prevention. National Healthcare Safety Network (NHSN). <https://www.cdc.gov/nhsn/about-nhsn/index.html> (2023).
39. Linkevicius, M. et al. Rapid cross-border emergence of NDM-5-producing *Escherichia coli* in the European Union/European economic area, 2012 to June 2022. *Euro Surveill.* **28**, 2300209 (2023).
40. Mujahid, F., Rasool, M. H., Shafiq, M., Aslam, B. & Khurshid, M. Emergence of carbapenem-resistant uropathogenic *Escherichia coli* (ST405 and ST167) strains carrying bla(CTX-M-15), bla(NDM-5) and diverse virulence factors in hospitalized patients. *Pathogens* **13**, 964 (2024).
41. Timm, M. R., Russell, S. K. & Hultgren, S. J. Urinary tract infections: pathogenesis, host susceptibility and emerging therapeutics. *Nat. Rev. Microbiol.* **23**, 72–86 (2025).

42. Perer, E. et al. Case report: Long-term follow-up of patients who received a FimCH vaccine for prevention of recurrent urinary tract infections caused by antibiotic resistant *Enterobacteriaceae*: a case report series. *Front Immunol.* **15**, 1359738 (2024).
43. Mikheenko, A., Pribelski, A., Saveliev, V., Antipov, D. & Gurevich, A. Versatile genome assembly evaluation with QUAST-LG. *Bioinformatics* **34**, i142–i150 (2018).
44. Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068–2069 (2014).
45. Minh, B. Q. et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020).
46. Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q. & Vinh, L. S. UFBoot2: Improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.* **35**, 518–522 (2017).
47. Letunic, I. & Bork, P. Interactive tree of life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. *Nucleic Acids Res.* **52**, W78–W82 (2024).
48. Sullivan, M. J., Petty, N. K. & Beatson, S. A. Easyfig: a genome comparison visualizer. *Bioinformatics* **27**, 1009–1010 (2011).
49. Ingle, D. J. et al. In silico serotyping of *E. coli* from short read data identifies limited novel O-loci but extensive diversity of O:H serotype combinations within and between pathogenic lineages. *Micro. Genom.* **2**, e000064 (2016).
50. Feldgarden, M. et al. AMRFinderPlus and the reference gene catalog facilitate examination of the genomic links among antimicrobial resistance, stress response, and virulence. *Sci. Rep.* **11**, 12728 (2021).
51. Wishart, D. S. et al. PHASTEST: faster than PHASTER, better than PHAST. *Nucleic Acids Res.* **51**, W443–w450 (2023).
52. Hadfield, J. et al. Phandango: an interactive viewer for bacterial population genomics. *Bioinformatics* **34**, 292–293 (2017).
53. Siguier, P., Perochon, J., Lestrade, L., Mahillon, J. & Chandler, M. ISfinder: the reference centre for bacterial insertion sequences. *Nucleic Acids Res.* **34**, D32–D36 (2006).
54. Phan, M.-D. et al. Plasmid-mediated ciprofloxacin resistance imparts a selective advantage on *Escherichia coli* ST131. *Antimicrobial Agents Chemother.* **66**, e02146–02121 (2022).
55. The European Committee on Antimicrobial Susceptibility Testing. *Breakpoint Tables for Interpretation of MICs and Zone Diameters*. <http://www.eucast.org> (2023).
56. Institute, C. L. S. *CLSI M100—Performance Standards for Antimicrobial Susceptibility Testing; 35th Edition*. <https://www.clsi.org/shop/standards/m100/> (2025).
57. Sarkar, S., Ulett, G. C., Totsika, M., Phan, M.-D. & Schembri, M. A. Role of capsule and O antigen in the virulence of uropathogenic *Escherichia coli*. *PLoS ONE* **9**, e94786 (2014).
58. Nhu, N. T. K. et al. A convergent evolutionary pathway attenuating cellulose production drives enhanced virulence of some bacteria. *Nat. Commun.* **15**, 1441 (2024).

Acknowledgements

This work was supported by an Australian National Health and Medical Research Council (NHMRC) Ideas grant to MAS, M-DP and NTKN

(2001431) and an NHMRC Ideas grant to MAS and M-DP (2037698). BMF is supported by a University of Queensland Health Research Accelerator grant. We thank the Centers for Disease Control and Prevention (CDC) National Healthcare Safety Network (NHSN), as well as other surveillance networks, for making genome sequence data used in this study publicly available.

Author contributions

Performed the experiments: L.L.W., B.P., Z.J.L., N.T.K.N., T.C., M.-D.P., K.M.P., K.A.R., C.S. Conceived and designed the experiments: L.L.W., M.-D.P., P.N.A.H., D.L.P., M.J.S., B.M.F., M.A.S. Analysed the data: L.L.W., M.-D.P., B.P., K.A.R., Z.J.L., N.T.K.N., B.M.F., M.A.S. Supervised aspects of the project: M.-D.P., N.P., T.J.K., P.N.A.H., D.L.P., M.J.S., B.M.F., M.A.S. Wrote the manuscript: L.L.W., M.-D.P., B.M.F., M.A.S. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no conflict of interest.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-025-63477-0>.

Correspondence and requests for materials should be addressed to Brian M. Forde or Mark A. Schembri.

Peer review information *Nature Communications* thanks Kate Baker, Henry Schreiber IV, and Kelly Wyres for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2025