

# Dominance of *Escherichia coli* sequence types ST73, ST95, ST127 and ST131 in Australian urine isolates: a genomic analysis of antimicrobial resistance and virulence linked to F plasmids

Dmitriy Li<sup>1,2</sup>, Paarthiphan Elankumaran<sup>1,2</sup>, Timothy Kudinha<sup>3</sup>, Amanda K. Kidsley<sup>4</sup>, Darren J. Trott<sup>4</sup>, Veronica Maria Jarocki<sup>1,2,\*</sup> and Steven Philip Djordjevic<sup>1,2,\*</sup>

## Abstract

Extraintestinal pathogenic *Escherichia coli* (ExPEC) are the most frequent cause of urinary tract infections (UTIs) globally. Most studies of clinical *E. coli* isolates are selected based on their antimicrobial resistance (AMR) phenotypes; however, this selection bias may not provide an accurate portrayal of which sequence types (STs) cause the most disease. Here, whole genome sequencing (WGS) was performed on 320 *E. coli* isolates from urine samples sourced from a regional hospital in Australia in 2006. Most isolates (91%) were sourced from patients with UTIs and were not selected based on any AMR phenotypes. No significant differences were observed in AMR and virulence genes profiles across age sex, and uro-clinical syndromes. While 88 STs were identified, ST73, ST95, ST127 and ST131 dominated. F virulence plasmids carrying *senB-cjrABC* (126/231; 55%) virulence genes were a feature of this collection. These *senB-cjrABC+* plasmids were split into two categories: pUTI89-like (F29:A-B10 and/or >95% identity to pUTI89) ( $n=73$ ) and non-pUTI89-like ( $n=53$ ). Compared to all other plasmid replicons, isolates with pUTI89-like plasmids carried fewer antibiotic resistance genes (ARGs), whilst isolates with *senB-cjrABC+*/non-pUTI89 plasmids had a significantly higher load of ARGs and class 1 integrons. F plasmids were not detected in 89 genomes, predominantly ST73. Our phylogenomic analyses identified closely related isolates from the same patient associated with different pathologies and evidence of strain-sharing events involving isolates sourced from companion and wild animals.

## DATA SUMMARY

The whole genome sequences of 320 isolates described here were uploaded to the Sequence Read Archive (SRA) in the National Centre for Biotechnology and Information (NCBI), under BioProject number PRJNA842786 with BioSample accession numbers SAMN28687489 to SAMN28687824. The whole genome sequences of 24 companion animal isolates were also uploaded to the SRA under BioProject number PRJNA858941. These 24 isolates were previously analysed in Saputra *et al.* (2017). Antimicrobial resistance in clinical *Escherichia coli* isolated from companion animals in Australia. *Vet Microbiol.*;211:43–50. doi: 10.1016/j.vetmic.2017.09.014.

Received 12 December 2022; Accepted 20 June 2023; Published 20 July 2023

**Author affiliations:** <sup>1</sup>Australian Institute for Microbiology & Infection, University of Technology Sydney, Ultimo, NSW, Australia; <sup>2</sup>Australian Centre for Genomic Epidemiological Microbiology, University of Technology Sydney, NSW, Australia; <sup>3</sup>Central West Pathology Laboratory, Charles Sturt University, Orange, NSW, Australia; <sup>4</sup>School of Animal and Veterinary Science, The University of Adelaide, Adelaide, South Australia, Australia.

**\*Correspondence:** Veronica Maria Jarocki, veronica.jarocki@uts.edu.au; Steven Philip Djordjevic, steven.djordjevic@uts.edu.au

**Keywords:** UTIs; ExPEC; F plasmids; pUTI89; *senB*; antimicrobial resistance.

**Abbreviations:** AMR, antimicrobial resistance; ARG, antimicrobial resistance gene; CARD, comprehensive antibiotic resistance database; ESBL, extended-spectrum beta lactamase; ExPEC, extraintestinal pathogenic *Escherichia coli*; HGT, horizontal gene transfer; HPI, high pathogenicity island; IS, insertion sequence; MDR, multi-drug resistance; MDS, multidimensional scaling; MLST, multilocus sequence type; NSW, New South Wales; OBH, Orange Base Hospital; PAI, pathogenicity associated island; pMLST, plasmid MLST; RST, replicon sequence type; SNP, single nucleotide polymorphism; ST, sequence type; UPEC, uropathogenic *E. coli*; UTI, urinary tract infection; VAG, virulence-associated gene; VFDB, virulence factor database; WGS, whole genome sequencing.

**Data statement:** All supporting data, code and protocols have been provided within the article or through supplementary data files. Additional files are available with the online version of this article.

001068 © 2023 The Authors



This is an open-access article distributed under the terms of the Creative Commons Attribution License.

### Impact Statement

Urinary tract infections (UTIs) are costly, reduce patient quality of life and can lead to the development of more life-threatening conditions like urosepsis. Extraintestinal pathogenic *Escherichia coli* (ExPEC) are the most common cause of UTIs and are often treated with antibiotics, which enhances antimicrobial resistance (AMR) evolution and selection. Genomic studies of retrospective collections of *E. coli* linked to UTI are uncommon. Knowledge of the genetic characteristics of *E. coli* that cause UTIs from retrospective collections provides insight into how AMR and virulence gene carriage changes with time and sheds light on the evolutionary mechanisms that underpin these events. AMR in Australia and worldwide is rising and the landscape of lineages present and dominating in the ExPEC population is constantly changing and influenced by geography. Therefore, it is important to apply modern technology to backdated clinical ExPEC populations to study how AMR evolves. Equally important is to identify and study local transmission events, like outbreaks, to provide better policies for prevention. This research could also provide insight into the genomic bases of uropathogenicity, and possibly future ways to counteract it. Most modern genomic epidemiological studies of ExPEC focus on antibiotic-resistant lineages, such as ST131, due to their enhanced clinical significance, but this leaves sequence types which are not as resistant poorly understood, obfuscating disease burden, limiting capacity for identification of emerging pathogens and limiting our understanding of the genomic context from which pandemic ExPEC lineages emerge.

## INTRODUCTION

Urinary tract infections (UTIs) incur an enormous cost burden to society and are the leading clinical presentation that drives antibiotic prescription [1]. In Australia, UTIs cause an estimated 69823 annual hospitalizations [2] and cost the nation's health system AU\$909 million annually [3]. Extraintestinal pathogenic *Escherichia coli* (ExPEC) are the leading cause of UTIs and are the most frequently isolated Gram-negative pathogen globally [4]. Additionally, ExPEC are responsible for bloodstream and wound infections, neonatal meningitis, and are the most frequent cause of ventilator-associated pneumonia [5]. Of more than 13000 *E. coli* sequence types (STs), only a small subset of pandemic lineages are responsible for the vast majority of ExPEC infections [6]. Pandemic ExPEC lineages, such as ST131, have also been central to a global increase in extended-spectrum  $\beta$ -lactamase (ESBL)-producing *E. coli*, as well as resistance to other clinically important antibiotic classes [4, 7, 8]. A combination of sulphonamide and trimethoprim is a standard treatment for UTIs but resistance rates to these frontline antibiotics are increasing globally, leading to the elevated use of extended-spectrum  $\beta$ -lactams and fluoroquinolones [9–13]. Previously we performed whole genome sequencing (WGS) on 67 trimethoprim-resistant ExPEC from patients attending a regional Australian hospital in 2006–2008 [14] and reported genes conferring resistance to extended-spectrum  $\beta$ -lactams, heavy metals and quaternary ammonium ions co-occurring with genes encoding resistance to trimethoprim.

The gut is a major reservoir of ExPEC, with vast numbers of these organisms shed into wastewater and diverse agricultural environments where they become exposed to frequent and often constant antimicrobial selection pressures, particularly in municipal wastewater, in food animal production and in animal faecal holding ponds. Gastrointestinal carriage of major ExPEC clonal lineages is influenced by frequent host-to-host transmission facilitated by sexual contact, international travel, contaminated food and water consumption, and interactions with wildlife, livestock and companion animals [15–19]. Constant recolonization with and across different hosts and repeated exposure to aquatic and terrestrial environments undoubtedly influence how *E. coli* acquires genetic information by horizontal gene transfer (HGT). Genetic features that enable colonization of different hosts, food, livestock, companion animals, wildlife and water sanitation practices profoundly influence pathogen biology, particularly regarding the evolution of emerging lineages, and is at the forefront of predictive infectious disease management. Much of the value inherent in studying successful pandemic ExPEC lineages lies in understanding what genetic features can be attributed to their global success, notwithstanding that biological success can evolve by diverse and often convergent paths. What is clear is that the acquisition of mobile genetic elements, particularly F plasmids [18, 20–22], phage [23] and genomic islands [24, 25], contributes to lineage evolution.

The association of F virulence plasmids, such as pUTI89 and ColV, in *E. coli* lineage evolution, host range and zoonosis is an increasingly important area of enquiry [18]. However, a detailed analysis linking the carriage of these plasmids in urinary tract isolates of *E. coli* has not been conducted. In pUTI89 and related plasmids, carriage of the *cjr* operon and the putative enterotoxin gene *senB* are considered important for virulence [26]. A study that interrogated a cohort of 34176 *E. coli* genome sequences (2570 STs) showed pUTI89 [replicon sequence type (RST) F29:A-B10] was overwhelmingly linked to *E. coli* sourced from humans but was almost entirely absent from 13027 *E. coli* isolates recovered from poultry, pigs and cattle [18]. F plasmids with RST F29:A-B10 have been associated with specific sublineages of major pandemic ExPEC lineages ST131, ST73, ST69 and ST95 [14, 18, 20]. In that same study, ColV-like plasmids were represented among *E. coli* sourced from poultry [2327/4254 (55%)] but also evident in 720/4425 (16%) human ExPEC isolates [18]. ColV virulence plasmids are found in *E. coli* that: (i) cause extraintestinal disease in humans and poultry as well as commensal *E. coli* [18, 27–29]; (ii) are required for avian pathogenic *E. coli* to cause colibacillosis

[30]; (iii) have been linked with zoonotic *E. coli* infections [18]; and (iv) display resistance to chlorine [31]. However, incomplete metadata linked with isolates deposited in public databases precludes a more thorough investigation of the association of these plasmids in *E. coli* linked to UTIs. This observation has clear implications in shedding light on a deeper understanding of the One Health aspects of *E. coli* disease.

Here we have undertaken a comprehensive phylogenomic analysis of 320 *E. coli* isolates from the urine of patients experiencing different clinical afflictions but predominantly with a uro-pathological focus (cystitis, pyelonephritis) from a single rural hospital in New South Wales (NSW), Australia.

## METHODS

### Sample collection

Urine samples from patients presenting at Orange Base Hospital were collected over a 6 month period (13 May to 12 November 2006). Specimens were included if they yielded a bacterial count of  $>10^8$  c.f.u.  $l^{-1}$  and cell count of  $>10^8 l^{-1}$  for white blood cells and  $<10^3 l^{-1}$  for epithelial cells. Specimens were excluded if patients had known diabetes mellitus, diarrhoea, received antibiotic therapy in the last month prior admission or were menstruating. Using these inclusion/exclusion criteria, a total of 353 samples [326 midstream urine (MSU); 27 catheter specimen urine (CSU)] were collected from 322 patients. Urine specimens were collected as previously described [14]. Briefly, each participating physician received a protocol for urine collection and the diagnostic criteria for classification of the uro-clinical syndrome. A diagnosis of cystitis or pyelonephritis required specific manifestations, as recorded by the treating medical practitioner. Cystitis-defining manifestations included dysuria, frequent urination and/or suprapubic tenderness, without fever or loin pain. Pyelonephritis-defining manifestations included urinary symptoms with a fever of  $\geq 38^\circ\text{C}$  and flank pain, with or without nausea/vomiting. Semi-quantitative cultures were performed on horse blood, MacConkey and chromogenic agars, followed by conventional biochemical tests. Isolates were stored in 50% (v/v) glycerol in trypticase soy broth at  $-70^\circ\text{C}$  until further use.

### Whole genome sequencing and genome assembly

DNA extraction and sequencing were performed as described previously [32]. Briefly, DNA was extracted using an ISOLATE II Genomic DNA (Bioline) kit following the manufacturer's standard protocol for bacterial cells, except for the final DNA elution step in which DNase- and RNase-free water were used. Library preparation was performed by the UTS Core Sequencing Facility at the University of Technology Sydney, using the adapted Nextera Flex library preparation kit process, Hackflex [33]. Sequencing was performed using an Illumina Novaseq S4 flow cell,  $2\times 150$  bp, at Novogene. The quality of reads was assessed using fastp (v0.20.1) [34]. Quality control was performed using assembly-stats v1.0.1 (output in Additional File 1A, available in the online version of this article). Genomes were excluded if total length was under 3.5 Mb or above 6.5 Mb. After quality control measures, the final collection consisted of 320 *E. coli* genomes derived from urine samples collected from 292 patients (multiple samples collected from 21 patients at different time points during the study timeline; however, any clonal isolates were excluded from statistical analyses; see below). Mean N50, number of contigs and read coverage was 253239, 153 and  $\times 68$ , respectively.

### Phylogeny and SNP analyses

Maximum-likelihood core genome phylogenies were built using IQtree2 (v2.0.3) [35] with extended model selection and tree reconstruction by best-fit model (-m MFP), 1000 bootstrap replicates (-bb 1000), using a core genome alignment generated by Roary (v3.13.0) [36] and the fast core gene alignment with MAFFT (-e --mafft) option. All core genome phylogenies were confirmed using both marker gene-based phylogenies produced by Phylosift using the default settings [37], and SNP-based phylogenies produced by snippy v4.3.6 (github.com/tseemann/snippy) with default settings, and recombination filtering was performed using Gubbins v2.3.4 [38] (github.com/sanger-pathogens/gubbins) with default settings. Phylogroups were determined by EzClermont [39]. SNP heatmaps were built by making pairwise SNP distance matrices of core genome alignments produced by prokka using snp-dists v0.6.3 with default settings (github.com/tseemann/snp-dists) and visualized using R v4.1.2.

### Gene screening

All isolates were annotated using the prokka pipeline [40] set for *E. coli* (--genus Escherichia --species coli), forced GenBank/ENA/DDJB compliance (--compliant), standard bacterial genetic code (translation table 11) (--gcode 11). All isolates were screened for antimicrobial resistance genes (ARGs), virulence-associated genes (VAGs), the presence of ColV markers, multilocus sequence types (MLSTs; Achtman 7 Gene) and plasmid MLST (pMLST) as previously described [18], using a pipeline available at github.com/maxlummins/pipelord2\_0. This pipeline incorporates assembly-stats v1.0.1 (github.com/sanger-pathogens/assembly-stats), fastp v0.20.1, kraken2 v2.2.1 [41], mlst v2.19.0 (github.com/tseemann/mlst) and pointfinder [42]. Assemblies with a genome size of  $<4.5$  Mb and  $>6.5$  Mb and N50  $>30000$  were excluded. Gene screening was performed using abricate v1.0.1 (github.com/tseemann/abricate) in conjunction with the following databases: VFDB [43], CARD [44], ISfinder (isfinder.biotoul.fr/) and PlasmidFinder [45]. Genes were marked as present when detected by abricate and filtered using a custom R script (github.com/

maxcummins/abricateR) to have 90% identity and 90% coverage. The presence and length of the class 1 integron integrase *intI1* was determined using BLAST+ v2.8.1 [46]. Contigs harbouring class 1 integrons were further analysed in SnapGene v4.1.9 (snapgene.com) and an integron map was drawn using graphics software Krita v4.4.1 (krita.org). Regarding specific IncF plasmid type groupings – as per modified Liu *et al.* [47] criteria [47] – isolates were marked as being ColV plasmid positive if they contained a gene from four or more of the following sets: (i) *cvaABC* and *cvi*, (ii) *iroBCDEN*, (iii) *iucABCD* and *iutA*, (iv) *etsABC*, (v) *ompT* and *hlyF*, and (vi) *sitABCD* [47], with  $\geq 95\%$  identity and  $\geq 95\%$  gene coverage. Isolates with either  $\geq 90\%$  length and identity of pUTI89 (NC\_007941.1) or having IncF RST F29:A-B10 were considered to be a pUTI89-like plasmid. Plasmid mapping was performed using a custom R script available at [github.com/maxcummins/plasmidmapR](https://github.com/maxcummins/plasmidmapR). The presence of PAI-I<sub>CFT073</sub> (NC\_004431-P1), PAI-II<sub>CFT073</sub> (NC\_004431-P2), PAI-III<sub>536</sub> (X16664), PAI-IV<sub>APEC-O1</sub> (NC\_008563) and PAI-V<sub>536</sub> (AJ617685) was scored as either present ( $\geq 90\%$  sequence identity and  $\geq 95\%$  sequence coverage), partial ( $\geq 90\%$  identity and  $< 95\%$  but  $> 50\%$  coverage) and absent ( $< 50\%$  coverage).

## Statistical analyses

To visualize the clustering of the isolates based on their virulence or resistance gene profile, gene presence/absence matrices were used in conjunction with classical (metric) multidimensional scaling (MDS) performed in R Studio using *cmdscale* and visualized using *ggplot2* v3.3.0 ([ggplot2.tidyverse.org/](https://ggplot2.tidyverse.org/)). To determine any statistically significant differences in gene presence-absence between UTI- and non-UTI-associated isolates, *Scoary* 1.6.16 [48] was used with `-no_pairwise` flag. Isolates from same patients with the same ST and identical ARG and VAG profiles ( $n=16$ ) were excluded from statistical analysis in sections Antimicrobial resistance and Virulence-associated genes below.

Statistical tests were performed in R v4.1.2, chi-square to determine the difference between plasmid types and *intI1* carriage performed using a standard `chisq.test`, while the Pairwise Wilcoxon test with Benjamini-Hochberg *P*-value correction for multiple testing was used to compare the number of ARGs between different plasmid type groups utilizing the `pairwise_wilcox_test` from the *rstatix* ([cran.r-project.org/package=rstatix](https://cran.r-project.org/package=rstatix)) package. All tests were checked for statistical power using the `pwr` v1.3-0 R package ([cran.r-project.org/package=pwr](https://cran.r-project.org/package=pwr)), and significance was reported if the power of the test was  $\geq 0.8$ .

## RESULTS

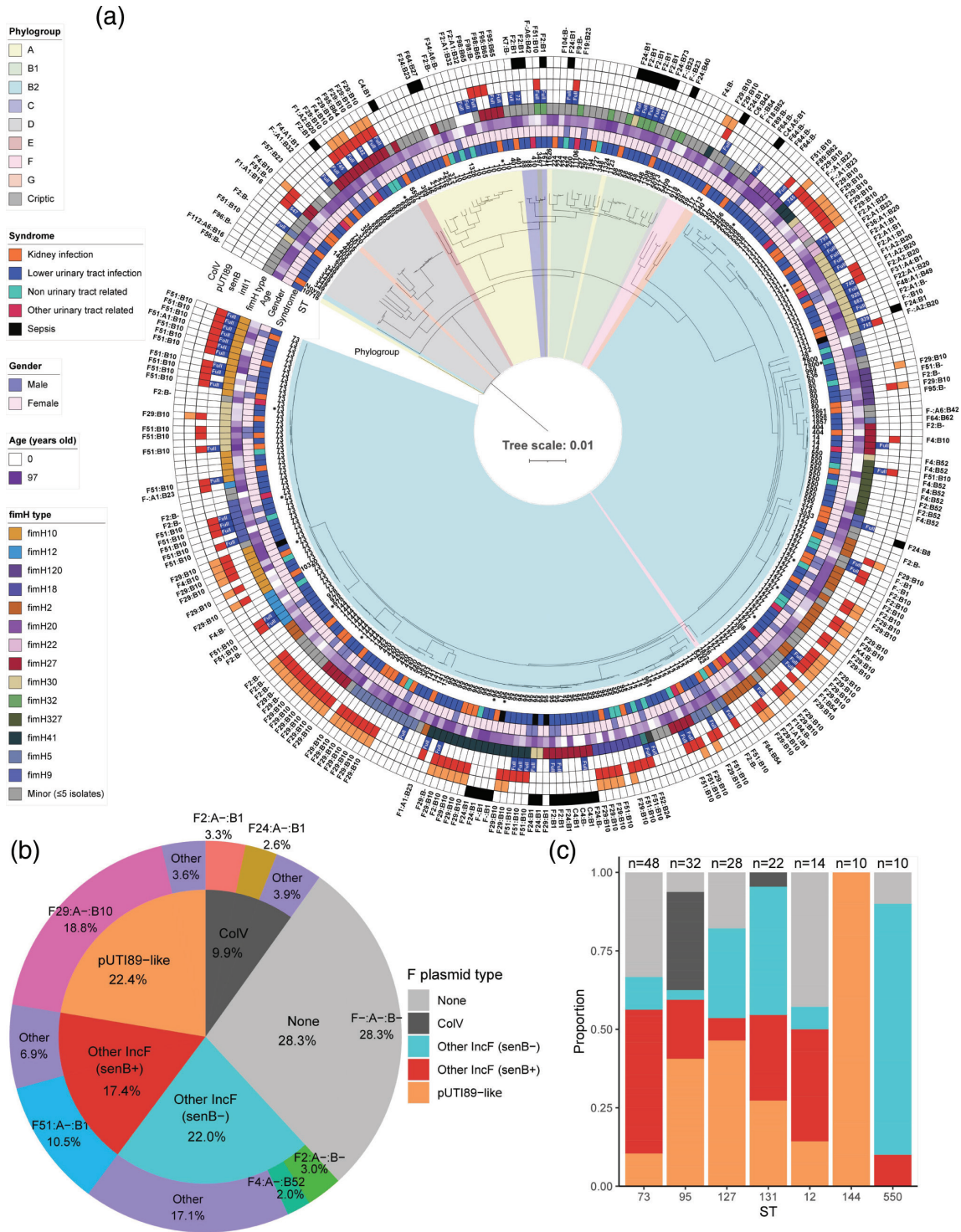
### Demography

This collection consisted of 320 *E. coli* draft genomes (mean contig size 46581 bp, mean genome size 5114541 bp) originating from isolates sourced from the urine of patients attending Orange Base Hospital (OBH), Australia, in 2006. Most isolates were sourced from patients with urinary tract disease, including kidney infections (20.6%,  $n=66$ ) and lower UTIs (64.7%,  $n=207$ ), while only five isolates (1.6%) were from patients with sepsis. Twenty-nine (9.1%) isolates were sourced from patients with diseases not related to the urinary tract. Isolates sourced from females dominated the collection ( $n=273$ ) compared to males ( $n=47$ ), with 11 patients pregnant at the time of sampling. Patient age varied from less than 1 month of age to 97 years old, with an average of 54.2 years. Fifteen (4.7%) isolates were acquired from patients  $\leq 3$  years old and 137 (42.8%) from patients  $\geq 65$  years old. Twenty-one patients were sampled more than once with an average isolation date difference of 38 days (range 0–156 days). Nineteen of these subsequent isolates were the same ST as the previous. Metadata for all OBH isolates is provided in Additional File 1B.

### Phylogeny

A maximum-likelihood core genome-based phylogenetic tree (Fig. 1a) saw the isolates cluster according to Clermont phylogroup [49]. EzClermont typing showed the presence of ten clades with isolates within phylogroup B2 dominating the collection: A ( $n=23$ , 7.2%), B1 ( $n=22$ , 6.9%), B2 ( $n=226$ , 70.6%), C ( $n=4$ , 1.3%), D ( $n=26$ , 8.1%), E ( $n=2$ , 0.6%), F ( $n=5$ , 1.6%), G ( $n=3$ , 0.9%), U ( $n=8$ , 2.5%) and cryptic ( $n=1$ , 0.3%). Eighty-eight STs were identified. The pandemic STs dominated the collection led by ST73 ( $n=51$ , 15.9%) followed by ST95 ( $n=34$ , 10.6%), ST127 ( $n=31$ , 9.7%), ST131 ( $n=24$ , 7.5%), ST12 ( $n=14$ , 4.4%) and ST144 ( $n=11$ , 3.4%). Isolates with ST10, ST69 and ST550 ( $n=10$ , 3.1% each), ST420 ( $n=7$ , 2.2% each), ST80 ( $n=6$ , 1.9%), ST349, ST538 and ST58 ( $n=5$ , 1.6% each) were also notable. The most common STs associated with specific uro-clinical syndromes are presented in Table 1. ST73 with serotype O6:H1 ( $n=19$ ) was the most prominent ST for patients diagnosed with lower UTIs. ST127 (O6:H31  $n=5$ ) was the most common ST in non-urinary tract-related diseases, and ST131 (serotype O25b:H4) and ST73 were frequently recovered from kidney infections. Three of the five isolates recovered from patients diagnosed with sepsis were ST95 and two carried a ColV F virulence plasmid.

F virulence plasmids that carry *cjrABC* together with *senB* and those that carry ColV virulence markers are associated with distinct pandemic ExPEC lineages [18, 21, 50, 51]. F virulence plasmids influence *E. coli* host range, zoonotic potential [18, 21, 32, 52] and AMR carriage in ExPEC [18, 50]. In this collection, 231 isolates carried F plasmids (72.2%), of which 126 isolates (54.5%) carried F plasmids with *senB-cjrABC*. These *senB-cjrABC*+ F plasmids consisted of 73 pUTI89-like plasmids (62 were F29:A-B10 and 11 were closely related to pUTI89 according to the selection criteria) and the remaining 53 were structurally different to pUTI89 [17 RSTs, predominantly F51:A-B10 ( $n=32$ )] (Fig. 1b). These data highlight the dominance of *E. coli* carrying F plasmids



**Fig. 1.** Orange Base Hospital isolate phylogeny and plasmid carriage. (a) Maximum-likelihood phylogenetic tree of 320 ExPEC isolates sourced from a rural hospital. Bars, the number of substitutions per site of alignment. Core genome SNP tree rooted on outgroup *Escherichia fergusonii* ATCC 35469<sup>T</sup> (not shown), including 2601 genes with a total alignment length of 2462224 bp, constructed by IQ-TREE2. Branch highlights coloured according to phylogroup identified by EzClermont, and leaf names represent the sequence type of each isolate. The outer ring represents the presence of the ColV plasmid according to Liu criteria [47] (black). Second outer ring: presence of pUTI89-like plasmid (orange). Third outer ring: presence of *senB* (red). Fourth outer ring: presence of *int1* (blue) and white text over the box represents truncation length of *int1*. F RSTs are displayed at the periphery. Isolates marked by an asterisk represent serial isolates with identical ARG and VAG profiles. (b) Pie chart showing the distribution of F plasmid types across the whole collection, with dark grey for ColV-positive isolates according to modified Liu *et al.* [47] criteria [47], orange for isolates positive for pUTI89-like plasmids, red for isolates positive for *senB* but negative for pUTI89-like plasmids, and blue for isolates positive for F plasmids but negative for *senB*, pUTI89 or ColV, as well as main *incF* RSTs present in these categories, with other for RSTs present in fewer than five isolates in particular categories. (c) Bar chart of proportions of the same plasmid type categories for STs with ten or more isolates.

**Table 1.** Most common sequence types (STs) and serotypes in isolates from different sources

Uro-clinical syndrome	Most common STs*	Most common serotypes
Kidney infections	ST73 (n=8), ST131 (n=7), ST95 (n=6), ST127 (n=5), ST144 and ST420 (n=4 each)	O25b:H4 (n=7), O134-Gp6:H31 (n=5), O6:H31 (n=5), O6:H1 and O1:H7 (n=4 each)
Lower urinary tract infections	ST73 (n=33), ST95 (n=21), ST127 (n=16), ST131 (n=12) and ST12/ST550 (n=7 each)	O6:H1 (n=18), O6:H31 (n=16), O1:H7 (n=10), O25b:H4 (n=9) and O75:H5 (n=8)
Non-urinary tract-related (e.g. myocardial infarction, pneumonia, migraine)	ST127, ST73 and ST12 (n=4 each), ST10 and ST95 (n=3 each)	O6:H31 (n=4)
Other urinary tract-related (e.g. renal failure, kidney stones, renal transplant)	ST127 (n=3), ST131 and 73 (n=2 each)	O6:H31 (n=3) and O25b:H4 (n=2)
Sepsis	ST95 (n=2), ST131 and ST73 (n=1 each)	O25b:H4, O1:H7, O2-O50-Gp7:H1 and O2-O50-Gp7:H4 (n=1 each)

\*No statistically significant associations between most common STs and uro-clinical syndromes found.

with *senB-cjrABC* virulence genes in urinary tract disease. Notably, *senB-cjrABC*-positive plasmids were present in all 11 ST144 isolates and most ST69 (n=10) and ST127 (n=17) isolates (Fig. 1c). ColV F virulence plasmids with diverse RSTs were detected in 32 (10%) genomes and were represented by 16 STs with ST95 the predominant lineage (n=12). F24:A-:B1 and F2:A-:B1 were the most common ColV RST (n=10 each).

*E. coli* with ST73, ST95, ST127 and ST131 accounted for nearly half (43.7%) of the collection. Additional core genome phylogenies were constructed for each of these four STs by including publicly available *E. coli* genomes with the same ST that originated from Australia. To provide a One Health perspective, genomes from non-human host origins were included and their relatedness to our clinical isolates was determined using SNP analyses.

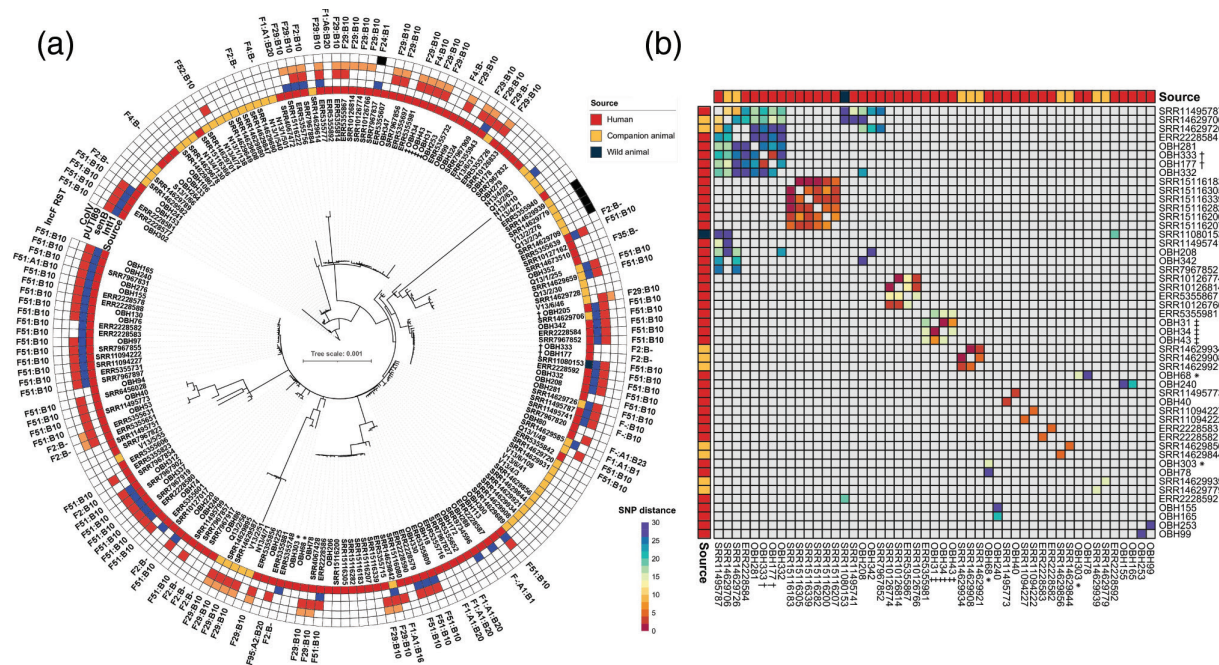
### ST73 phylogeny

ST73 is a serologically and phylogenetically diverse ST [53]. ST73 isolates were the most abundant ST from the OBH collection. A separate core genome maximum-likelihood phylogenetic tree was reconstructed using ST73 isolates from this collection, all ST73 genomes available on Enterobase from Australia (human n=84, wild animal n=1 and companion animal n=30, date range: 2006–2019; metadata in Additional File 1C), and 24 ST73 genomes from companion animals as described previously [54, 55] (Fig. 2a). This phylogenetic analysis showed that companion animal-sourced isolates tended to cluster together and that these isolates were less likely to carry F plasmid replicons (i.e. 27% of isolates originating from humans were F plasmid negative, as opposed to 67% in isolates from animals). Isolates from OBH were broadly distributed with no evident clustering by source or serotype (Fig. 2a). Nevertheless, we did identify multiple instances of small SNP differences in isolates originating from different hospitals as well as between clinical isolates and those derived from companion animals (Fig. 2b). Specifically, we highlighted clusters of ST73 isolates that displayed very close SNP distances: (i) an isolate retrieved from a wild animal (flying fox, SRR11080153) which differed by 27 SNPs to companion animal-sourced isolate SRR14629706 – the flying fox-sourced isolate also differed by 29 and 20 SNPs respectively to human-sourced isolates SRR11495787 and ERR2228592; (ii) two closely related companion animal-sourced isolates SRR14629706 and SRR14629726 to seven human-sourced isolates (average 22 SNPs, range 10–33); and (iii) closely related human-sourced isolates, including four OBH isolates from different patients comprising two with lower UTI, one with renal calculi and one with acute myocardial infarct, with an average of 24 SNPs across them (range 3–29 SNPs).

We were able to identify F plasmid replicon STs in 116 (61%) ST73 isolates depicted in Fig. 2. ST73 was dominated by F plasmids carrying *senB-cjrABC* (n=85, 73.3%) with only five isolates (4.3%) predicted to carry a ColV plasmid including a single ST73 carrying F24:A-:B1 and a cluster of four phylogenetically related isolates with an F untypable replicon ST. pUTI89 (F29:A-:B10 n=27) ST73 isolates were dispersed across the phylogeny. F plasmids related to pUTI89 comprised F2:A-:B- (n=2), F2:A-:B10 and F4:A-:B10 (both n=1). *senB-cjrABC* plasmids distinct from pUTI89 with F51:A-:B10 (50/85; 58.8%) was the dominant RST. A significant number (n=74; 38.9%) of ST73 genomes did not carry an F plasmid (F-:A-:B-). Liu et al. [47] developed criteria that have been useful in identifying *E. coli* carrying ColV plasmids [21, 47]. We noted that 98 ST73 isolates tested false positive using these ColV marker criteria many of which (n=55, 56%) were F-:A-:B- by pMLST. Read mapping analyses (data not shown) failed to provide evidence of the presence of ColV plasmids in these isolates. Rather these false-positive ColV ST73 isolates carried a suite of genomic islands (see ‘Virulence-associated genes’ section below) noted for carriage of ColV virulence markers. To address this, we increased the stringency of the Liu et al. [47] criteria [47] to ≥95% identity and ≥95% gene coverage as outlined in the Methods and this removed the false positive status of these ST73 isolates.

### ST95 phylogeny

A core genome maximum-likelihood phylogenetic tree of 34 ST95 isolates from the OBH as well as an additional 51 genomes from Enterobase derived from Australian ST95 sourced from poultry (n=25), humans (n=22) and the environment



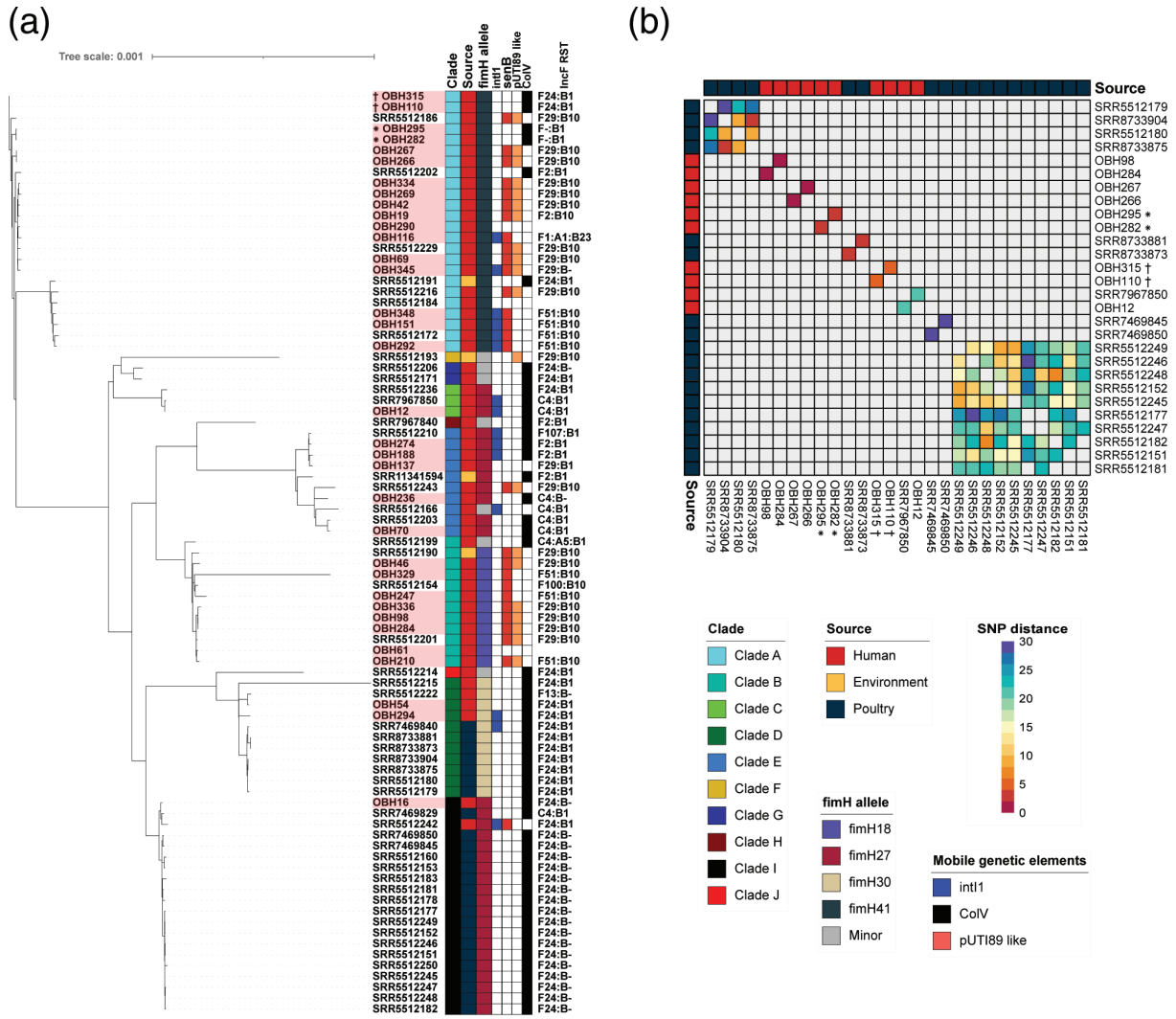
**Fig. 2.** Phylogeny and SNP analysis of ST73 isolates. (a) Mid-point rooted core genome-based maximum-likelihood phylogenetic tree of ST73 isolates (alignment length 3201980bp). Leaf names correspond to isolate names; isolate names highlighted by red colour are from this study. Inner ring represents the source of the isolates, and second ring represents the presence of *intI1* (blue). Third ring indicates the presence of *senB* (red) with the fourth ring indicating the presence of the pUTI89-like plasmid (orange). The outer ring represents the presence of the ColV plasmid according to modified Liu criteria [47] (black). Outside of the rings, F RSTs are shown at the periphery. (b) Pairwise SNP distance heatmap showing isolates with 30 SNPs (purple) or <30 SNPs (purple to red colour scale). Values higher than the threshold or self-comparisons are greyed out. Serial isolates are marked by \*, † or ‡.

( $n=4$ ) (date range: 2001–2019; metadata in Additional File 1C) with varying Hierarchical Clustering (HC20) designations was constructed (Fig. 3a). Previously we showed that *E. coli* ST95 isolates can be categorized into ten distinct clades (A–J) [18]. A maximum-likelihood core genome phylogenetic tree (Fig. 3a) revealed that the majority of ST95 isolates from the OBH collection belong to the human-centric clade A ( $n=17$ , 50%), followed by clade B ( $n=8$ , 24%) and clade E ( $n=5$ , 15%). No OBH isolates were identified as belonging to clades F, H or J. Small SNP differences between isolates from different sources were only observed between one human isolate and a poultry isolate (both in Clade I; 64 SNPs) (Fig. 3b). However, examples of close human-sourced ST95 isolates from clade A were found including isolates from the same patients: OBH266 (cystitis) and OBH267 (cystitis) which were collected on the same day differed by 2 SNPs; OBH282 (nephritis) and OBH295 (sepsis) were 3 SNPs apart; and OBH110 (cystitis) with OBH315 (cystitis) which were collected on the same day differed by 6 SNPs. Also present were closely related isolates from different patients: OBH284 (chronic cystitis) and OBH98 (urethritis) were 2 SNPs apart; and OBH12 (urethritis) and SRR7967850 (sepsis) were 21 SNPs apart. Notably, isolate SRR7967850 was sourced from Concord Repatriation Hospital in Sydney in 2013, approximately 7 years and over 200 km apart.

pUTI89-like plasmids were present in 13 isolates (38%), most of which belonged to clade A ( $n=8$ , 61%) with *fimH41* and serotype O1:H7. Similarly, ColV plasmids were present in 12 isolates (35%), but as expected most ColV-positive isolates did not belong to human-dominated clades A or B ( $n=8$ , 67%) but were dominant in clades I and in clades C and E. The distribution of ColV and pUTI89-like plasmids across the major STs identified in this study can be viewed in Fig. 1(c). Notably, ST95 had the highest carriage of ColV plasmids, while other STs including ST127, ST69 and ST144 were predominantly associated with pUTI89-like plasmid carriage. Consistent with this we recently described the carriage of pUTI89-like plasmids in ST127 from canine and human origins [32].

### ST127 phylogeny

A core genome maximum-likelihood phylogenetic tree of ST127 was built using an additional 41 Australian ST127 isolates from Enterobase (human  $n=18$ , companion animal  $n=22$ , wild animal  $n=1$ ; date range: 2006–2019, metadata in Additional File 1C) (Fig. 4a). The 31 ST127 isolates from this collection were broadly distributed among four main clusters. Overall, ST127 isolates were found to be diverse, with 6178 SNPs separating the two most distant isolates (companion animal-sourced isolates SRR14629755 and SRR14629640). However, we did identify several closely related isolates (Fig. 4b) including: (i) four isolates,



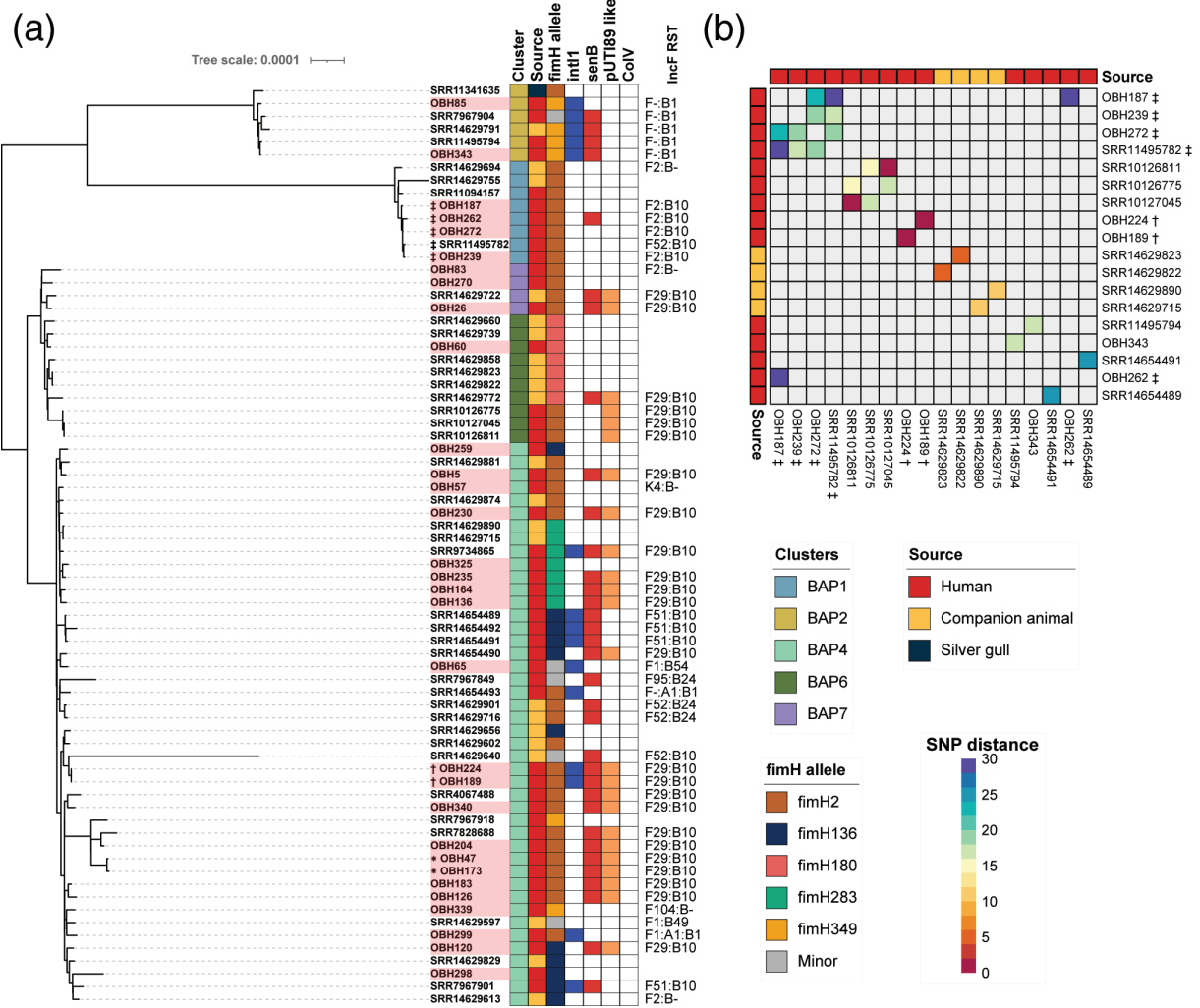
**Fig. 3.** Phylogeny and SNP analysis of ST95 isolates. (a) Mid-point rooted core genome-based maximum-likelihood phylogenetic tree of ST95 isolates from this collection, plus 51 additional Australian ST95 isolates of varying HC20. Total alignment length: 3 448 827 bp. Leaf names correspond to isolate names; isolates from this collection are highlighted in red. Coloured squares on the right represent metadata for each isolate, first column represents ST95 clades, second refers to source, third column represents *fimH* allele, fourth presence of *int11* (blue), fifth presence of *senB* (red), and sixth presence of pUTI89-like plasmid (orange). Outer ring represents presence of the ColV plasmid according to Liu criteria [47] (black). Outside label corresponds to F RST of the isolate. (b) Pairwise SNP distance heatmap showing pairs with 30 SNPs (purple) or lower (purple to red gradient); values higher than the threshold or self-comparisons are greyed out. Serial isolates are marked by \* or †.

all originating from one patient of OBH (OBH187, 239, 272, SRR11495782) experiencing different uro-clinical syndromes (pyelonephritis, cystitis and glomerulonephritis) over a 2 month period, had an average of 24 SNPs (range 18–33 SNPs); (ii) two clinical isolates from different patients (OBH343, cystitis; and SRR11495794, urosepsis), were found to have distance of 18 SNPs; and (iii) two isolates, OBH189 and OBH224, isolated 24 days apart from the same patient with a recurrent UTI were indistinguishable (0 SNPs).

**ST131 phylogeny**

We recently performed comprehensive genomic analyses on all available *E. coli* ST131 genomes originating from Australia ( $n=284$ ) [56], so here we provide an SNP-based recombination filtered maximum-likelihood phylogenetic tree of ST131 isolates of this collection plus eight Australian ST131 isolates, from 2008 to 2019 (metadata in Additional File 1C), representing the major clades C1, C2, B and A [56, 57] (Fig. 5). Most isolates from this collection belonged to clade C ( $n=14$ , 58%), followed by seven (29%) in clade B and three (13%) in clade A. The diversity of F RSTs differed among ST131 clades: OBH isolates in clades A and B carried three different F RSTs, F29:B10 ( $n=6$ ), F2:A1:B23 ( $n=2$ ) and F-:A1:B23 ( $n=2$ ), and all of these were pUTI89-like plasmids. Clade





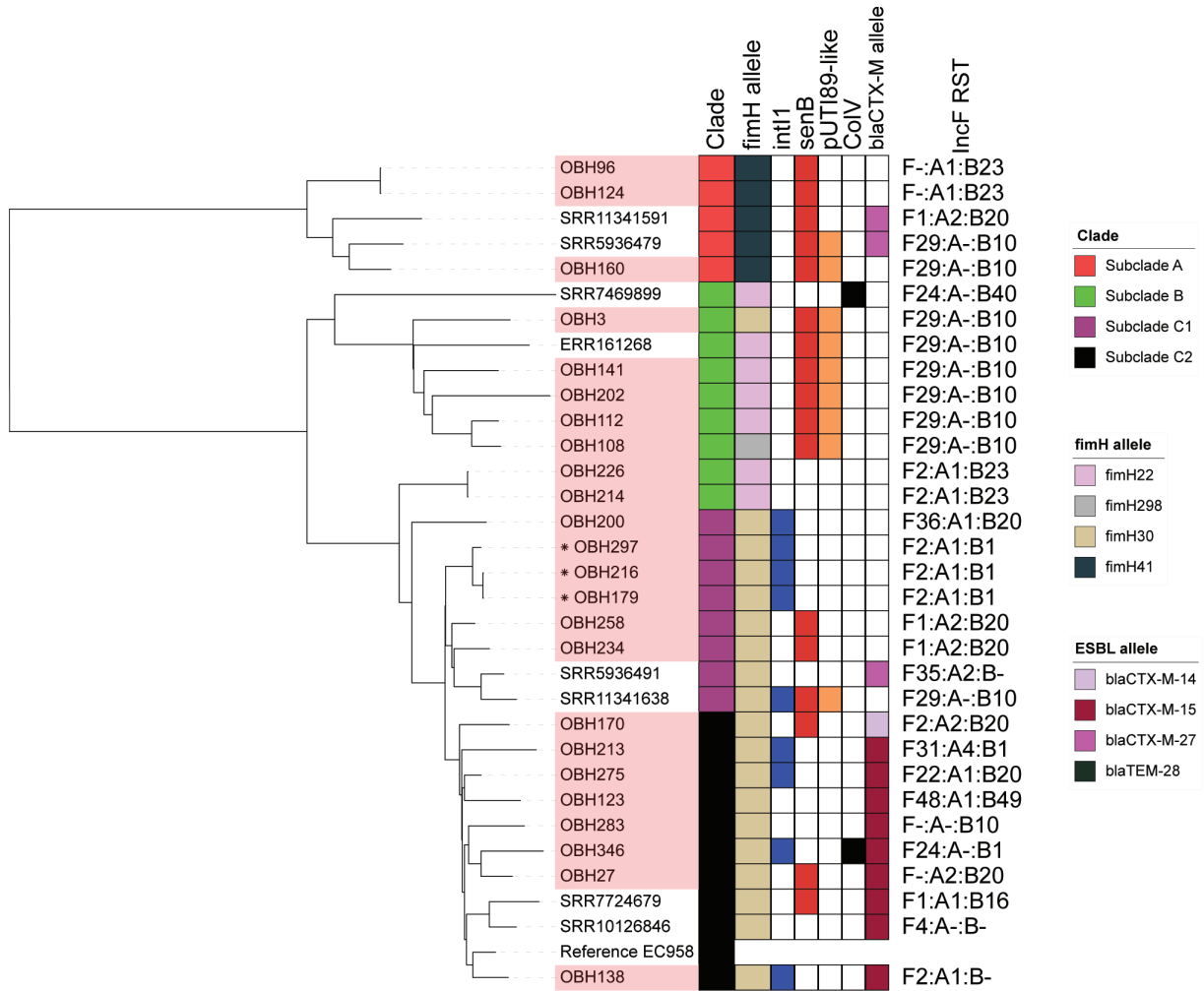
**Fig. 4.** Phylogeny and SNP analysis of ST127 isolates. (a) Mid-point rooted core genome-based maximum-likelihood phylogenetic tree of ST127 isolates. Rooted by an ST372 isolate (SRA accession number: SRX10969655; not shown on the tree). Total alignment length of 3 579 832 bp. Leaf names correspond to isolate names, with isolates from this study highlighted in red. Coloured squares on the right represent metadata for isolate: first column represents source, second column *fimH*, third presence of *int1* (blue), fourth presence of pUTI89 like plasmid (orange) and fifth presence of *senB* (red). Outer ring represents presence of the ColV according to Liu criteria [47] (black). Outside label corresponds to *incF* RST of the isolate. (b) Pairwise SNP distance heatmap showing pairs with 30 SNPs (purple) or <30 (purple to red colour gradient); values higher than the threshold or self-comparisons are greyed out. Serial isolates are marked by \*, † or ‡.

C1 isolates also had three types present (F2:A1:B1 *n*=3, F1:A2:B20 *n*=2, F36:A1:B20 *n*=1) while in clade C2 every isolate (*n*=8) had a unique F plasmid RST.

ST131 isolates are known to be highly clonal, especially within the C1 and C2 clades [56], but using a 30 SNP cut-off we only found three isolates within this threshold: OBH179 (pyelonephritis), OBH216 (glomerulonephritis) and OBH297 (cystitis). These were all sourced from the same patient. OBH216 was isolated 27 days after OH179 and differed by 28 SNPs and OBH297 was isolated 15 days after OBH216 and differed by 12 SNPs. Only one ST131 isolate from our collection was found to be ColV positive (OBH346, clade C2).

**Antimicrobial resistance**

We identified 95 ARGs in this collection (Fig. 6a; Additional File 1D). Only 12 isolates (4.3%) carried genes encoding ESBLs, and these were predominantly found in ST131 isolates (*n*=9). Only isolates in ST131 clade C carried ESBLs with seven positive for *bla*<sub>CTX-M-15</sub>, and one for *bla*<sub>CTX-M-14</sub> and *bla*<sub>TEM-28</sub> each. Sulfonamide resistance genes were found in more than a third of all isolates (37.2%; *n*=113), and the most common gene was *sul1* (*n*=66), followed by *sul2* (*n*=59) and *sul3* (*n*=3). Similarly, genes conferring resistance to aminoglycosides were identified in 32.2% of isolates (*n*=98), the three most common being *aph*(3'')-Ib (*strA*) (*n*=51),

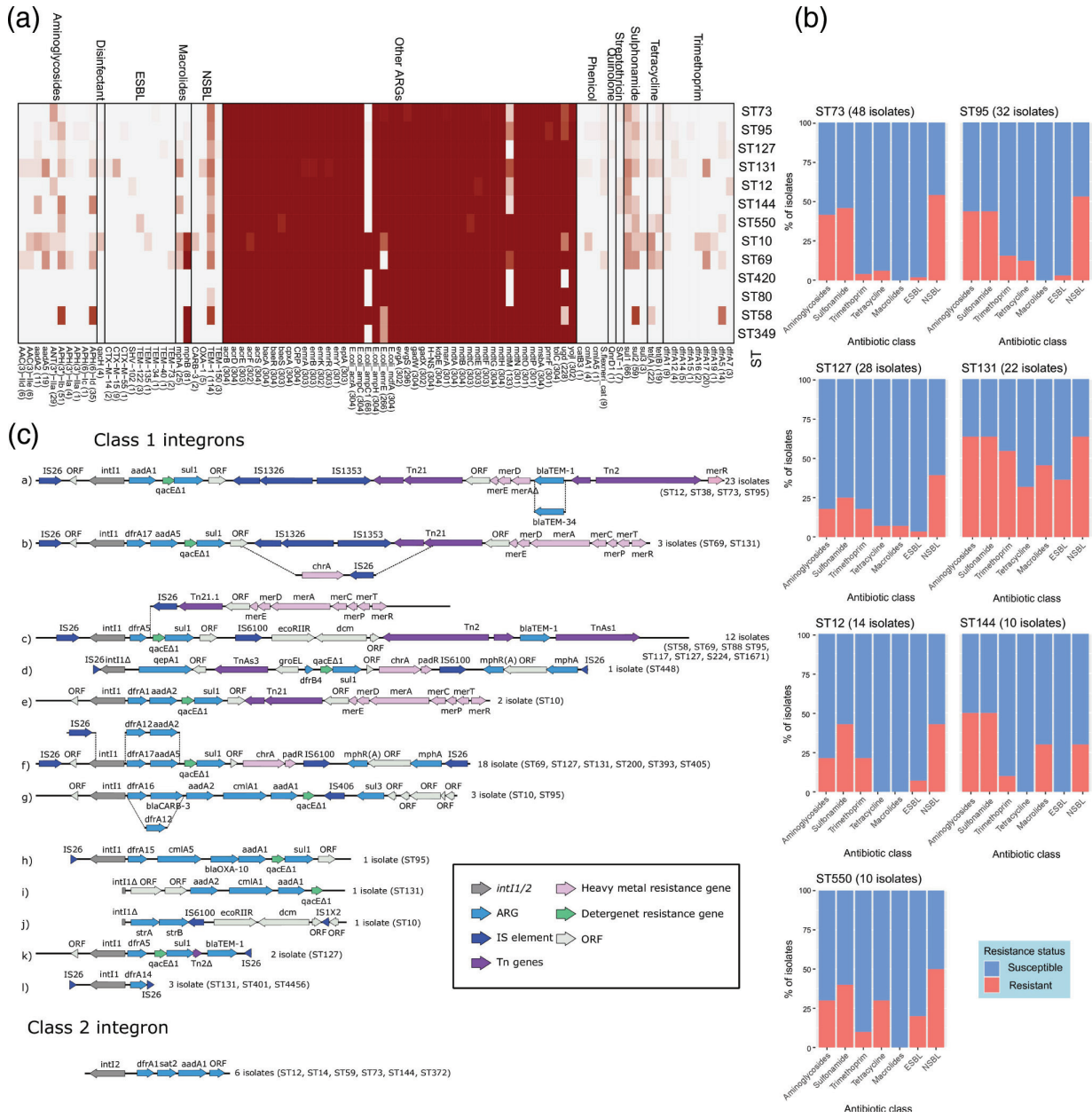


**Fig. 5.** Phylogeny and SNP analysis of ST131 isolates. Mid-point rooted core genome-based maximum-likelihood phylogenetic tree of ST131 isolates. Total alignment length of 3 702 378bp. Leaf names correspond to isolate names; isolates from the current study are highlighted in red. Coloured squares on the right represent metadata for isolates; first column: clade, second column: *fimH* alleles, third: *int11* (blue), fourth: presence of pUT189 like plasmid (orange), fifth: presence of *senB* (red). Outer column represents the presence of the ColV plasmid according to Liu criteria [47] (black). Outside label corresponds to incF RST. \*Serial isolates.

*aph(6)-Id (strB)* ( $n=35$ ; additional 16 truncated) and *ant(3'')-IIa* ( $n=29$ ). Genes conferring resistance to trimethoprim were found in approximately one-fifth of this collection (19%;  $n=58$ ), the most prevalent being *dfrA17* ( $n=20$ ), *dfrA5* ( $n=14$ ) and *dfrA1* ( $n=9$ ). The recently characterized *dfrB4* gene [58] was identified in just one isolate (ST448 OBH273). Tetracycline resistance genes were identified in 41 (13.5%) isolates, with *tet(A)* ( $n=22$ ) being most common, followed by *tet(B)* ( $n=19$ ) and *tet(D)* ( $n=1$ ). Only one isolate carried both *tet(A)* and *tet(B)*. Genes conferring resistance to macrolides, chloramphenicol, streptothricin, quinolones and fosfomycin were less common. The macrolide resistance gene *mphA* was found in 8.2% ( $n=25$ ) of isolates. Genes conferring chloramphenicol resistance were found in seven isolates (2%) (*cmlA1*:  $n=5$ , *cmlA5*:  $n=1$ , *catB3*:  $n=1$ ). The streptothricin resistance gene *sat2* was found in seven isolates (two ST73, one of ST12, ST14, ST59, ST144 and ST372 each). The quinolone resistance gene *qnrD1* was identified in only one ST127 isolate. Similarly, just one ST1727 isolate carried the fosfomycin resistance gene *fosA7*.

The collection was also screened for specific point mutations in *gyrA* and *parC* genes that confer resistance to fluoroquinolones. The *gyrA*-1AB mutation was identified in 4.9% of the isolates ( $n=15$ ), most commonly in ST131 ( $n=11$ ), but also in ST393 ( $n=2$ ), ST448 and ST405 ( $n=1$  each). The *gyrA*-1AB and *parC*-1aAB dual mutations were identified in 3.9% of isolates ( $n=12$ ): 11 in ST131 (Clades C1 and C2) isolates and in one ST405 isolate.

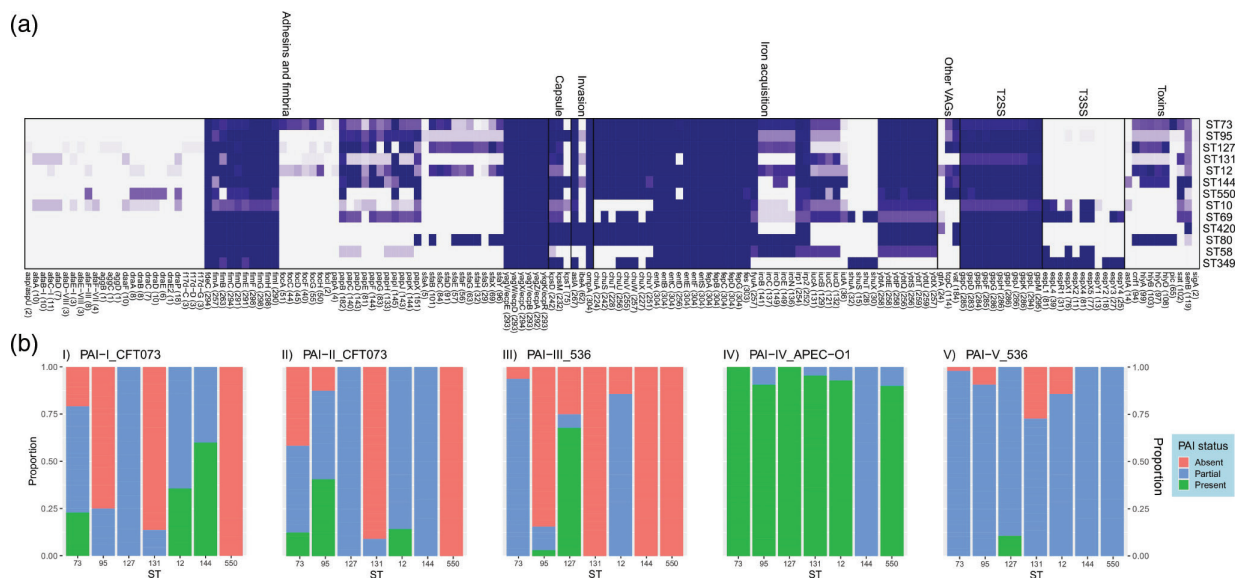
Overall, the carriage of ARGs was distributed as follows: 3.9% of isolates for gentamycin resistance (based on *aac(3)-IId*, *aac(3)-IIe*, *aph(3')-IIa* [59]), 37.5% for sulfonamide resistance alone, 18.1% for trimethoprim and sulfonamide resistance combined,



**Fig. 6.** Genomic antibiotic resistance analysis. (a) Gene screening heat map for antibiotic resistance genes for STs with  $\geq 5$  isolates. (b) Bar chart of genomic antibiotic resistance in STs with  $\geq 5$  isolates. ESBL – extended-spectrum  $\beta$ -lactamase, NSBL – narrow-spectrum  $\beta$ -lactamase. (c) Class 1 (a–l) and class 2 integron structures identified in this collection. Integron integrase genes in grey, antimicrobial resistance genes (ARG) in cyan, insertable (IS) elements in blue, transposon (Tn) genes in purple, heavy metal resistance genes in pink, detergent resistance genes in green, and other ORF in white.

4.3% for ESBLs, and 3.9% for fluoroquinolone resistance. Sixty-one per cent ( $n=185$ ) of isolates did not carry genes or mutations conferring resistance to any of these antibiotic classes and no isolates carried carbapenemase genes.

Despite the skewed frequency of ARGs overall, there were no associations between uro-clinical syndrome and ARG profiles nor any statistically significant differences between uro-clinical syndrome and the mean number of ARGs. Similarly, there were no associations between ARG profiles or the mean number of ARGs in terms of patients’ biological sex and age (Additional Files 1E and 1F). The distribution of ARGs conferring resistance to clinically important antibiotics differed among the major STs in this collection, with ST131, ST69 and ST10 isolates exhibiting robust multidrug-resistant (MDR) profiles (Fig. 6b).



**Fig. 7.** Genomic virulence analysis. (a) Gene screening heat map for virulence-associated genes for STs with  $\geq 5$  isolates. T2SS – type 2 secretion system, T3SS – type 3 secretion system. (b) Presence status of five pathogenicity-associated islands (PAIs) for STs with  $\geq 10$  isolates.

Class 1 integrons can capture and express ARGs and lead to an MDR phenotype. The capture and shuffling of ARGs are mediated via the integrase *intI1* [60]. Here, a total of 80 isolates (26.3%) contained *intI1*. Most ( $n=67$ ) carried the non-truncated form of *intI1*. In the remaining 13 isolates, seven different *intI1* truncations were observed (Fig. 1a), the most prevalent being *intI1* <sub>$\Delta 745$</sub>  ( $n=4$ ). Overall, the most common class 1 integron structure was *intI1-aadA1-qacE $\Delta$ 1-sul1* associated with a Tn21 transposon, and a Tn2 transposon (containing the *bla*<sub>TEM-1</sub> gene) inserted into the *mer* operon (mercury resistance) [Fig. 6c(a)]. This particular class 1 integron confers resistance to aminoglycosides, sulphonamides, narrow-spectrum beta-lactams and quaternary ammonium compounds. This structure was dominant in ST73 ( $n=16/23$ ) and isolates from patients with lower UTIs ( $n=17/23$ ). In one isolate (ST73), an SNP in the *bla*<sub>TEM-1</sub> gene gave rise to *bla*<sub>TEM-34'</sub>, thereby providing putative ESBL resistance. Only one isolate (an ST131 sourced from a kidney infection) was identified carrying two class 1 integrons [Fig. 6c(i), c(l)]. The highest prevalence of class 1 integrons was found in ST131 (54.5%,  $n=12$ ) followed by ST10 (44.4%,  $n=4$ ) and ST12 (42.8%,  $n=6$ ), which correlates with these STs having MDR profiles as mentioned above. There was a significant difference found between plasmid type and class 1 integron carriage ( $\chi^2=83.56$ ,  $df=4$ ,  $P<2.2e-16$ ), with *senB-cjrABC+*/non-pUTI89-like plasmids (encompassing 17 IncF RSTs; most common F51:A-B10) being most associated with *intI1*. In particular, 77% ( $n=27$ ) of isolates carrying F51:A-B10 were *intI1* positive (Additional File 1D). Conversely, isolates that carried pUTI89-like plasmids had significantly lower *intI1* carriage ( $n=5$ , 7.9% vs  $n=78$ , 46.6%,  $\chi^2=15$ ,  $df=1$ ,  $P=0.0001$ ). Isolates that carried *senB-cjrABC*/non-pUTI89-like plasmids also carried on average the most ARGs associated with HGT ( $n=4.5$ ), which was significantly different to other plasmid types (pUTI89-like  $P=1.48e-07$ ; *senB*-IncF plasmids  $P=0.000162$ ; no IncF plasmid  $P=1.83e-16$ ) (Additional File 1F).

The class 2 integron *intI2-dfrA1-sat2-aadA1* was identified in six isolates (Fig. 6c), providing resistance to trimethoprim, streptomycin and aminoglycosides. Other integron classes were not detected.

### Virulence-associated genes

A total of 218 VAGs were identified in the collection (Fig. 7a; Additional File 1D). The molecular definition of uropathogenic *E. coli* (UPEC; a subset of ExPEC) is defined as the presence of  $\geq 3$  of the following genes: *chuA*, *fyuA*, *ycfV* and *vat* [61]. By this definition, 219 (72%) isolates qualified as UPEC, but 75 isolates obtained from patients with UTIs did not possess  $\geq 3$  of these genes, including 60 isolates from lower UTIs and 15 from kidney infections. Notably, a scoary analysis found no statistically significant difference in VAG carriage (or any other gene) between UTI- and non-UTI-associated isolates (Additional File 1G). Similarly, no associations between the overall VAG profile and uro-clinical syndrome, sex and age were observed (Additional Files 1E and 1F). Across all age groups, sexes and pathologies, genes involved in iron acquisition, adhesion, immune evasion and toxins were widespread. Iron acquisition systems play an important role in surviving the iron-scarce environment of the urinary tract [62]. The ferric yersiniabactin uptake receptor *fyuA* was found in 85.3% ( $n=273$ ) of isolates. The iron-regulatory proteins *irp1* ( $n=270$ , 84.3%) and *irp2* ( $n=268$ , 83.7%) were similarly common, with the complete yersiniabactin siderophore operon *ybtAEPQSTYX* present in 85% ( $n=272$ ) of isolates. The Salmochelin siderophore system encoded by *iroBCDEN* was identified in 45.6% ( $n=146$ ) of isolates. The full aerobactin operons *iucABCD* and *iutA* were present in 11.6% ( $n=37$ ) of isolates.

Specificity for uroepithelium has been demonstrated for type 1 fimbriae and P-fimbriae which are important for UTI-causing ExPEC [62]. Type I fimbriae genes were present in most isolates and *fimH*, which encodes a protein that binds to uroepithelial associated  $\alpha$ -D-mannosylated proteins, was present in 98.1% of isolates ( $n=314$ ). P-fimbriae, encoded by the *pap* operon, bind to receptors located in the upper urinary tract and are associated with nephritis [63]. However, we did not find a higher association between the presence of *pap* operon genes and kidney infections.

ExPEC have well-documented immune evasion mechanisms and produce toxins facilitating damage of the host tissues [64, 65]. Immune evasion mechanisms include increased serum survival protein (*iss*) present here in 74.7% ( $n=239$ ) of isolates, VirB5-like protein TraT (*traT*) observed in 65.6% of isolates ( $n=210$ ) and outer-membrane protein T (*ompT*) detected in 10% of isolates ( $n=32$ ). Toxins identified included enterotoxin TieB protein (*senB*) observed in 38.8% ( $n=124$ ) of isolates, full alpha-haemolysin operon (*hlyABCD*) present in 31.9% ( $n=102$ ) of isolates, cytotoxic necrotizing factor 1 (*cnf1*) found in 30.9% of isolates ( $n=99$ ) and heat-stable enterotoxin 1 (*astA*) identified in 4.4% ( $n=14$ ) of isolates.

Iron acquisition systems, adhesins and toxins are commonly found on pathogenicity-associated islands (PAIs). We screened for the presence of five PAIs (Additional File 1h) – PAI-I<sub>CFT073</sub> carrying the haemolysin operon (*hlyABCD*), PAI-II<sub>CFT073</sub> with P fimbriae operon (*papACDEFGHIJK*), PAI-III<sub>536</sub> containing adhesion (*sfhABCDEFGHIJSX*) and the iron acquisition genes (*iroBCDEN*, putative haemin receptor), PAI-IV<sub>APEC-01</sub> which carries the yersiniabactin operon (*fyuA*, *irp1*, *irp2*, *ybtAEPQSTUX*), and PAI-V<sub>536</sub> containing adhesins (*pixABCDFGHJ*), *pgtABCP*) and capsule-related operons (*kpsCEFMSTU*).

The presence of ExPEC-associated PAIs was scored as either present ( $\geq 90\%$  sequence identity and  $\geq 95\%$  sequence coverage), partial ( $\geq 90\%$  identity and  $< 95\%$  but  $> 50\%$  coverage) or absent ( $< 50\%$  coverage). Evidence for a complete PAI-I<sub>CFT073</sub> was found in 24 isolates, mostly in ST73 ( $n=11$ ) and ST144 ( $n=6$ ) isolates. A partial copy of PAI-I<sub>CFT073</sub> was found in 99 isolates. Most isolates (60.6%;  $n=60$ ) carrying a partial copy of PAI-I<sub>CFT073</sub> contained the full *hly* operon but not a full *pap* operon. PAI-II<sub>CFT073</sub> was identified in 22 isolates, mostly with ST95 ( $n=13$ ), followed by ST73 ( $n=6$ ). Nearly half (45.8%;  $n=22$ ) of ST73, 47% ( $n=15$ ) of ST95, 85.7% ( $n=12$ ) of ST12, and all ST127 and ST144 isolates carried a partial island, with most missing *ireA* (63.3%,  $n=69$ ) and *papA* (70.6%,  $n=77$ ). PAI-III<sub>536</sub> was present in 21 isolates, almost exclusively within ST127 isolates ( $n=19$ ), while ST73 contained the most (93.7%;  $n=45$ ) partial versions of this PAI followed by ST12 (85.7%;  $n=12$ ). Isolates carrying a partial PAI-III<sub>536</sub> were missing the putative lysin/cadaverine transporter and the putative cadaverine decarboxylase genes. PAI-IV<sub>APEC-01</sub> was present in most isolates ( $n=186$ ; 61.2%). The first 28321 bp (36.5%) of this PAI has 97.92% sequence identity to the high pathogenicity island (HPI). Most phylogroup B2 isolates carried a full PAI-IV<sub>APEC-01</sub> ( $n=170$ ) or a partial version ( $n=39$ ). Of the 39 partial versions, all encompassed a complete HPI. Only three isolates contained a complete PAI-V<sub>536</sub>, all of which were ST127, but only a small proportion (14.2%,  $n=30$ ) of phylogroup B2 isolates did not possess a partial PAI-V<sub>536</sub> (all PAI maps and mosaic plots showing cross-sectional distributions of PAIs across sex, age and pathologies can be viewed in Additional File 1h).

In summary, the top six pandemic STs (ST73, ST131, ST95, ST69, ST127 and ST12) often carried the five PAIs to different degrees (Fig. 7b). PAI-IV<sub>APEC-01</sub> was present in almost all B2 phylogroup isolates. Out of all the STs, ST127 isolates had the highest proportion of carrying either complete or partial versions of all five PAIs.

## DISCUSSION

ExPEC comprise diverse STs; however, globally only a subset is responsible for most infections. A recent systematic review found that 85% of ExPEC infections were attributed to just 20 STs, the top five being ST131 > ST69 > ST10 > ST405 > ST38 [6]. In this study of urine-sourced clinical *E. coli* isolates ( $n=320$ ) from an Australian regional hospital in 2006, the most common STs were ST73 > ST95 > ST127 > ST131 > ST12 which accounted for 48% of all isolates. Most of the 20 STs flagged by Manges *et al.* (2019) [6] were represented in this collection, with the exception of ST648, ST354, ST167, ST617, ST23 and ST1193. *E. coli* ST1193 is a rapidly emerging global MDR lineage that is particularly prevalent in Australia [51]. Its notable absence in our 2006 collection suggests that ST1193 had not taken hold in ExPEC that cause UTI in Australia at this time. While ExPEC STs vary in abundance and diversity within human populations [66], our observed difference between the top ST distributions probably stemmed from the fact that most studies included in the review were biased because the collections were based on antibiotic resistance [6]. Indeed, our previous study on trimethoprim-resistant *E. coli* UTI isolates from the same hospital [14] found that the most common STs were ST131 and ST69. However, when sample selection was not based on AMR phenotypes, our current results reflect other studies from the UK, USA and Canada, wherein ST73, ST95, ST127 and ST131 were the most frequently isolated STs from large cohorts of UTIs [67–69].

Given the importance of AMR, ongoing surveillance of resistant populations undoubtedly has merit, but it does not provide an accurate portrayal of the STs most responsible for UTIs. *E. coli* ST73 as a primary causative agent of UTIs is likely to be underrepresented in the literature as the lineage is often described as pan-sensitive to antibiotics [70, 71] and plasmid naïve [72]. This was not so in this collection, as many ST73 isolates carried ARGs conferring resistance to  $\beta$ -lactams, sulphonamides, aminoglycosides, class 1 integron integrases *intI1*, and at least one plasmid replicon. Notably, however, only one ST73 isolate carried an ESBL resistance gene. The most common plasmid type detected in ST73 isolates was

*senB-cjrABC+*/non-pUTI89-like IncF plasmids with RST F51:A-B10. We found this plasmid type was significantly associated with higher ARG carriage. F51:A-B10 plasmids are known to frequently carry class 1 integrons with *aadA1* [73]. Indeed, the most common class 1 integron structure in this collection was *intI1-aadA1-qacEΔ1-sul1*. This integron is not only found in *E. coli* but also in numerous other bacterial species including *Aeromonas*, *Bacillus*, *Citrobacter*, *Klebsiella*, *Pseudomonas*, *Salmonella* and *Vibrio* [74]. While we found F51:A-B10 to be associated with ARGs, pUTI89-like plasmids (also *senB-cjrABC+*), which were dominated by F29:A-B10 in our collection, were associated with lower ARG carriage. This is somewhat unsurprising given that this plasmid lineage has been reported to be associated with pan-susceptible ExPEC strains [50].

Despite the positive association of *senB-cjrABC+*/non-pUTI89-like IncF plasmids and ARGs, the overall ARG carriage in this retrospective collection was mostly low, particularly regarding ARGs conferring resistance to antibiotics commonly prescribed for complicated UTIs such as gentamycin, ESBLs and fluoroquinolones. However, ARGs encoding resistance to trimethoprim/sulphonamide were considerable. The majority of ESBL genes were found in ST131, which is unsurprising given that this lineage is thought to have played a central role in the global increase of ESBL-producing *Enterobacteriales* [75]. Though now dominating in Australian ST131 C1 and A clades [56], *bla*<sub>CTX-M-27</sub> genes were not detected in any of the ST131 isolates from OBH, indicating that this gene may not have yet reached this population. Most of the genotypic fluoroquinolone resistance we observed was attributed to dual *parC* and *gyrA* mutations in ST131 isolates. Interestingly, however, while dual *gyrA*-1AB and *parC*-1aAB were previously thought to be highly specific to ST131 clade C isolates [76], we found these specific mutations in an ST405 isolate. Recently we also reported these mutations in ST131 clade A isolates [56]. Overall, ST131 isolates possessed the most robust MDR profiles, followed by ST10 and ST69 isolates. These STs had the highest prevalence of *intI1*, giving credence to class 1 integrons as reliable markers for MDR [77]. Curiously, while truncated *intI1* genes are common among *E. coli* populations [78] and can be used as epidemiological markers [79], here only 16% of *intI1* genes had deletions, which may speak to the retrospective nature of this collection, or possibly that truncated *intI1* genes are more prevalent in non-human isolates. For example, in an Australian study of 425 critically drug-resistant *E. coli* from gulls sampled in 2012, 242 (57%) were determined to carry class 1 integrons. Of these, 64% showed 3'-truncations in *intI1*, most often associated with IS26, with identical truncations found across multiple lineages. Often these truncations are missed by high-throughput gene identification.

Recent studies have shown that rural populations are more likely to receive inappropriate antibiotics for inappropriate durations than urban populations [80, 81]. OBH services regional, rural and remote communities, so one could speculate that the carriage of ARGs and class 1 integrons in UTI isolates would be higher here than in an urban hospital setting. However, the overall *intI1* carriage was lower compared to that reported in a study of UTI isolates from three metropolitan Australian hospitals during the same time period (26.3% vs 34%) [82]. One possible explanation for this is that urban surface waters and sediments can have higher ARG loads and plasmid carriage compared to rural samples [83], meaning that the urban environment could be a significant driver for AMR.

UTIs are predominately community-acquired and, beyond the human gastrointestinal tract, the environment, sewage and abattoir waste, retail meats, and animals (wild, food and companion) have all been suggested as ExPEC reservoirs [18, 21, 27, 28, 32, 47, 52, 79]. To explore potential interspecies movement, we performed phylogenetic and SNP analyses on our top four STs (ST73, ST95, ST131, ST127) and compared them to human and non-human-sourced *E. coli* genomes of Australian origin. There is currently no universal SNP threshold to infer *E. coli* transmission, though recent efforts have determined that a 17 SNP cut-off is useful in determining nosocomial outbreaks [84, 85]. Given that we were interested in identifying potential cross-sectoral species transmission, we used a putative SNP threshold of 30. ST73 was previously described as a human-specific ExPEC lineage [86, 87]. However, recent studies have shown that ST73 is prominent in cats [55, 88, 89], dogs [52, 54, 90], killer whales [91] and some avian populations [92]. Consistent with previous studies, our phylogenetic analysis showed ST73 as diverse [53, 93], and that animal-sourced isolates tended to cluster together [54, 55]. Nevertheless, we did identify clusters of very closely related (<30 SNPs) ST73 isolates from human and non-human sources. These low interspecies SNP counts are particularly relevant given the phylogenetic diversity of ST73. We previously showed the potential for interspecies movement in an Australian ST131 population [56], but ST131s have a relatively conserved core genome even amongst the different clades. Conversely, ST73 isolates can differ by thousands of SNPs [93] and indeed we found that some human-sourced ST73 isolates differed from other ST73 human-sourced isolates by up to 5952 SNPs. *E. coli* ST127 has been described as an emerging, highly virulent, human pathogen [67, 94] but has also been isolated from companion animals [32, 52, 54, 55, 88], killer whales [91] and bats [95]. Our investigation did not find any human-sourced ST127 isolates with <30 SNPs from isolates of animal sources. Nevertheless, a recent study of ST127 reported that some isolates from geographically proximal and distal human- vs. companion animal-sources varied by <30 SNPs [32]. *E. coli* ST95 is a prominent cause of both human and avian diseases [96], and the zoonotic potential of this lineage is well documented, particularly in relation to O1:H7 strains [97–99], or those belonging to clade I [18]. The most common serotype in this hospital ST95 population was O1:H7, but these were dispersed predominately among human-dominated clade A, and we found no human-sourced isolate under 30 SNPs from any non-human-sourced isolate.

It is evident that certain STs are more likely to cause UTIs, but the mechanisms behind pathogenesis remain elusive. Congruent with previous studies we found that ST73 was most commonly isolated from patients with cystitis [86], ST95 was associated with sepsis [100] and ST131 was most commonly retrieved from kidney infections [101]. However, we found no significant difference in VAG carriage between isolates from different pathologies. Indeed, we found no significant differences in VAGs, or any other gene, between UTI isolates and non-UTI isolates. Furthermore, we found several examples of isolates with small SNP distances in isolates from patients experiencing different pathologies. Even the molecular definition of UPEC is problematic. By definition, 72% of our isolates qualified as UPEC but excluded were 75 isolates from patients with UTIs. The majority of ExPEC VAGs encompass siderophore systems, urinary tract-specific adhesins and immune evasion effectors [102], all of which were ubiquitous in our collection. It is important to note that *E. coli* belonging to phylogroup B2, particularly ST73, ST131, ST95 and ST141, are dominant in the gut of healthy humans and their frequency has increased in the commensal faecal population over the past 40 years [103, 104]. These genes prime ExPEC for survival outside of the gastrointestinal tract but also play a role in its persistence in the human gut, while pathogenesis is likely to involve underlying patient factors, such as comorbidities and age [105], or a combination of VAGs. On that note, PAIs can encode several different VAGs and are known to play a major role in the evolution of ExPEC [106]. We observed a correlation between five specific PAIs and STs and found that overall ST127 isolates carried the most PAIs, a feature that was noted in a recent report of ST127 [32].

Our study has some limitations, namely the lack of temporal association between the human clinical and animal isolates, given that the OBH collection is from 2006 and the animal isolates used for comparison were all isolated later (2009–2019). Additionally, AMR phenotypes were not available for the isolates, but ARG presence and resistance phenotypes are typically highly congruent in *E. coli* [107].

## CONCLUSIONS

This study describes a phylogenomic analysis of a large cohort of *E. coli* from the urine of patients with uro-centric disease from a major rural hospital in NSW, Australia. While it is unsurprising that the top six STs are the pandemic lineages ST73 > ST95 > ST127 > ST131 > ST12 which accounted for 48% of all isolates, our study overlays the distribution of F virulence plasmids, PAIs and ARG cargo and provides deeper insights into lineage evolution. The collection is dominated by *E. coli* that carry plasmids with *senB-cjrABC* with those carrying F29:A-B10 (pUTI89) lacking carriage of antibiotic resistance whereas those carrying *senB-cjrABC* but not closely related to pUTI89 carry a greater antibiotic resistance gene load and were typically *intI1+*. We were able to provide evidence of isolate movements between patients within and across hospital settings, demonstrate the persistence of a clonal lineage from the same patient over short and considerable time periods, and importantly demonstrate evidence of occasional interspecies transmission of ST73, particularly between humans and companion animals. Our study underlines the importance of taking a One Health genomic approach to pathogen surveillance.

### Funding information

This work was partially funded by the Australian Centre for Genomic Epidemiological Microbiology (Ausgem), a collaborative partnership between the NSW Department of Primary Industries and the University of Technology Sydney. This project was also partially funded by an Australian Research Council Linkage grant LP130100736 with Zoetis Australia as the main industry partner.

### Acknowledgements

We would like to thank UTS sequencing facility, Dr Kay Anantanawat from the University of Technology Sydney, and Professor Sam Abraham from Murdoch University for their help with sequencing hospital and companion animal isolates.

### Author contributions

D.L.: analysis and interpretation of data, drafted and revised the work; P.E.: acquisition of data; T.K.: acquisition of data; A.K.K.: acquisition of data and revised the work; D.J.T.: conception and design of work, acquisition of data and revised the work; V.M.J.: conception and design of work, interpretation of data, drafted and revised the work; S.P.D.: conception and design of work, acquisition of data, drafted and revised the work.

### Conflicts of interest

The authors declare that there are no conflicts of interest.

### Ethical statement

The project was approved by Charles Sturt University and Sydney West Area Health Service research ethics committees. Since clinical information for patients with UTIs was provided anonymously by clinicians, patient consent was not required

### References

1. Abbo LM, Hooton TM. Antimicrobial stewardship and urinary tract infections. *Antibiotics* 2014;3:174–192.
2. AURA. Third Australian report on antimicrobial use and resistance in human health.: Australian Commission on Safety and Quality in Health Care (ACSQHC); 2019. <https://www.safetyand-quality.gov.au/sites/default/files/2019-06/AURA-2019-Report.pdf>
3. McCall B. Artificial intelligence bridges the complexity of intractable medical problems. *Lancet Digit Health* 2021;3:e8–e9.
4. 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;213:6–13.
5. Manges AR. *Escherichia coli* causing bloodstream and other extraintestinal infections: tracking the next pandemic. *Lancet Infect Dis* 2019;19:1269–1270.

6. Manges AR, Geum HM, Guo A, Edens TJ, Fibke CD, et al. Global Extraintestinal Pathogenic *Escherichia coli* (ExPEC) lineages. *Clin Microbiol Rev* 2019;32:e00135-18.
7. Zilberberg MD, Shorr AF. Secular trends in gram-negative resistance among urinary tract infection hospitalizations in the United States, 2000-2009. *Infect Control Hosp Epidemiol* 2013;34:940-946.
8. Bevan ER, Jones AM, Hawkey PM. Global epidemiology of CTX-M  $\beta$ -lactamases: temporal and geographical shifts in genotype. *J Antimicrob Chemother* 2017;72:2145-2155.
9. Wesolek JL, Wu JY, Smalley CM, Wang L, Campbell MJ. Risk factors for trimethoprim and sulfamethoxazole-resistant *Escherichia coli* in ED patients with urinary tract infections. *Am J Emerg Med* 2022;56:178-182.
10. Carmona-Cartaya Y, Hidalgo-Benito M, Borges-Mateus LM, Pereda-Novales N, González-Molina MK, et al. Community-acquired uropathogenic *Escherichia coli*, antimicrobial susceptibility, and extended-spectrum beta-lactamase detection. *MEDICC Rev* 2022;24:20-25.
11. García-Meniño I, Lumbreras P, Lestón L, Álvarez-Álvarez M, García V, et al. Occurrence and genomic characterization of clone ST1193 clonotype 14-64 in uncomplicated urinary tract infections caused by *Escherichia coli* in Spain. *Microbiol Spectr* 2022;10:e0004122.
12. Bean DC, Livermore DM, Papa I, Hall LMC. Resistance among *Escherichia coli* to sulphonamides and other antimicrobials now little used in man. *J Antimicrob Chemother* 2005;56:962-964.
13. Tano ZN, Kobayashi RK, Candido EP, Dias JB, Perugini LF, et al. Susceptibility to first choice antimicrobial treatment for urinary tract infections to *Escherichia coli* isolates from women urine samples in community South Brazil. *Braz J Infect Dis* 2022;26:102366.
14. Li D, Reid CJ, Kudinha T, Jarocki VM, Djordjevic SP. Genomic analysis of trimethoprim-resistant extraintestinal pathogenic *Escherichia coli* and recurrent urinary tract infections. *Microb Genom* 2020;6:mgen000475.
15. Manges AR, Johnson JR. Reservoirs of extraintestinal pathogenic *Escherichia coli*. *Microbiol Spectr* 2015;3.
16. Nesporova K, Wyrsh ER, Valcek A, Bitar I, Chaw K, et al. *Escherichia coli* sequence type 457 is an emerging extended-spectrum- $\beta$ -lactam-resistant lineage with reservoirs in wild-life and food-producing animals. *Antimicrob Agents Chemother* 2020;65:e01118-20.
17. Tarabai H, Wyrsh ER, Bitar I, Dolejska M, Djordjevic SP. Epidemic HI2 plasmids mobilising the carbapenemase gene *bla*<sub>IMP-4</sub> in Australian clinical samples identified in multiple sublineages of *Escherichia coli* ST216 colonising silver gulls. *Microorganisms* 2021;9:567.
18. Cummins ML, Reid CJ, Djordjevic SP. F plasmid lineages in *Escherichia coli* ST95: implications for host range, antibiotic resistance, and zoonoses. *mSystems* 2022;7:e0021022.
19. Kantele A, Lääveri T, Mero S, Häkkinen IMK, Kirveskari J, et al. Despite predominance of uropathogenic/extraintestinal pathotypes among travel-acquired extended-spectrum  $\beta$ -lactamase-producing *Escherichia coli*, the most commonly associated clinical manifestation is travelers' diarrhea. *Clin Infect Dis* 2020;70:210-218.
20. Stephens C, Arismendi T, Wright M, Hartman A, Gonzalez A, et al. F plasmids are the major carriers of antibiotic resistance genes in human-associated commensal *Escherichia coli*. *mSphere* 2020;5:e00709-20.
21. Reid CJ, Cummins ML, Börjesson S, Brouwer MSM, Hasman H, et al. A role for ColV plasmids in the evolution of pathogenic *Escherichia coli* ST58. *Nat Commun* 2022;13:683.
22. Johnson TJ, Enekave E, Miller EA, Munoz-Aguayo J, Flores Figueroa C, et al. Phylogenomic analysis of extraintestinal pathogenic *Escherichia coli* sequence type 1193, an emerging multidrug-resistant clonal group. *Antimicrob Agents Chemother* 2019;63:e01913-18.
23. Venturini C, Zingali T, Wyrsh ER, Bowring B, Iredell J, et al. Diversity of P1 phage-like elements in multidrug resistant *Escherichia coli*. *Sci Rep* 2019;9:18861.
24. Dobrindt U. (Patho-)genomics of *Escherichia coli*. *Int J Med Microbiol* 2005;295:357-371.
25. Galardini M, Clermont O, Baron A, Busby B, Dion S, et al. Major role of iron uptake systems in the intrinsic extra-intestinal virulence of the genus *Escherichia* revealed by a genome-wide association study. *PLoS Genet* 2020;16:e1009065.
26. Cusumano CK, Hung CS, Chen SL, Hultgren SJ. Virulence plasmid harbored by uropathogenic *Escherichia coli* functions in acute stages of pathogenesis. *Infect Immun* 2010;78:1457-1467.
27. 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;5:e000295.
28. Cummins ML, Reid CJ, Roy Chowdhury P, Bushell RN, Esbert N, et al. Whole genome sequence analysis of Australian avian pathogenic *Escherichia coli* that carry the class 1 integrase gene. *Microb Genom* 2019;5:e000250.
29. Moran RA, Holt KE, Hall RM. pCERC3 from a commensal ST95 *Escherichia coli*: a ColV virulence-multiresistance plasmid carrying a *sul3*-associated class 1 integron. *Plasmid* 2016;84-85:11-19.
30. Nolan LK, Barnes HJ, Vaillancourt J-P, Abdul-Aziz T, Logue CM. Colibacillosis. In: *Diseases of Poultry*. 2013. pp. 751-805.
31. Hicks SJ, Rowbury RJ. Virulence plasmid-associated adhesion of *Escherichia coli* and its significance for chlorine resistance. *J Appl Bacteriol* 1986;61:209-218.
32. Elankumaran P, Browning GF, Marends MS, Reid CJ, Djordjevic SP. Close genetic linkage between human and companion animal extraintestinal pathogenic *Escherichia coli* ST127. *Curr Res Microb Sci* 2022;3:100106.
33. Gaio D, Anantanawat K, To J, Liu M, Monahan L, et al. Hackflex: low-cost, high-throughput, Illumina Nextera Flex library construction. *Microb Genom* 2022;8:000744.
34. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 2018;34:i884-i890.
35. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol* 2020;37:1530-1534.
36. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, et al. Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics* 2015;31:3691-3693.
37. Darling AE, Jospin G, Lowe E, Matsen FA 4th, Bik HM, et al. PhyloSift: phylogenetic analysis of genomes and metagenomes. *PeerJ* 2014;2:e243.
38. Croucher NJ, Page AJ, Connor TR, Delaney AJ, Keane JA, et al. Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. *Nucleic Acids Res* 2015;43:e15.
39. Waters NR, Abram F, Brennan F, Holmes A, Pritchard L. Easy phylotyping of *Escherichia coli* via the EzClermont web app and command-line tool. *Access Microbiol* 2020;2:acmi000143.
40. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 2014;30:2068-2069.
41. Wood DE, Salzberg SL. Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biol* 2014;15:R46.
42. Zankari E, Allesøe R, Joensen KG, Cavaco LM, Lund O, et al. Point-Finder: a novel web tool for WGS-based detection of antimicrobial resistance associated with chromosomal point mutations in bacterial pathogens. *J Antimicrob Chemother* 2017;72:2764-2768.
43. Liu B, Zheng D, Zhou S, Chen L, Yang J. VFDB 2022: a general classification scheme for bacterial virulence factors. *Nucleic Acids Res* 2022;50:D912-D917.
44. Alcock BP, Raphenya AR, Lau TTY, Tsang KK, Bouchard M, et al. CARD 2020: antibiotic resistome surveillance



- with the comprehensive antibiotic resistance database. *Nucleic Acids Res* 2020;48:D517–D525.
45. Carattoli A, Zankari E, García-Fernández A, Voldby Larsen M, Lund O, et al. In Silico detection and typing of plasmids using plasmidFinder and plasmid multilocus sequence typing. *Antimicrob Agents Chemother* 2014;58:3895–3903.
  46. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, et al. BLAST+: architecture and applications. *BMC Bioinformatics* 2009;10:421.
  47. Liu CM, Stegger M, Aziz M, Johnson TJ, Waits K, et al. *Escherichia coli* St131-H22 as a foodborne uropathogen. *mBio* 2018;9:e00470–18.
  48. Brynildsrud O, Bohlin J, Scheffer L, Eldholm V. Rapid scoring of genes in microbial pan-genome-wide association studies with Scoary. *Genome Biol* 2016;17.
  49. Gonzalez-Alba JM, Baquero F, Cantón R, Galán JC. Stratified reconstruction of ancestral *Escherichia coli* diversification. *BMC Genomics* 2019;20:936.
  50. Stephens CM, Adams-Sapper S, Sekhon M, Johnson JR, Riley LW. Genomic analysis of factors associated with low prevalence of antibiotic resistance in extraintestinal pathogenic *Escherichia coli* sequence type 95 strains. *mSphere* 2017;2:e00390–16.
  51. Wyrsh ER, Bushell RN, Marena MS, Browning GF, Djordjevic SP, et al. Global phylogeny and F virulence plasmid carriage in pandemic *Escherichia coli* ST1193. *Microbiol Spectr* 2022;10:e0255422.
  52. Elankumaran P, Cummins ML, Browning GF, Marena MS, Reid CJ, et al. Genomic and temporal trends in canine ExPEC reflect those of human ExPEC. *Microbiol Spectr* 2022;10:e0129122.
  53. Bogema DR, McKinnon J, Liu M, Hitchick N, Miller N, et al. Whole-genome analysis of extraintestinal *Escherichia coli* sequence type 73 from a single hospital over a 2 year period identified different circulating clonal groups. *Microb Genom* 2020;6:e000255.
  54. Kidsley AK, O'Dea M, Saputra S, Jordan D, Johnson JR, et al. Genomic analysis of phylogenetic group B2 extraintestinal pathogenic *E. coli* causing infections in dogs in Australia. *Vet Microbiol* 2020;248:108783.
  55. Kidsley AK, O'Dea M, Ebrahimie E, Mohammadi-Dehcheshmeh M, Saputra S, et al. Genomic analysis of fluoroquinolone-susceptible phylogenetic group B2 extraintestinal pathogenic *Escherichia coli* causing infections in cats. *Vet Microbiol* 2020;245:108685.
  56. Li D, Wyrsh ER, Elankumaran P, Dolejska M, Marena MS, et al. Genomic comparisons of *Escherichia coli* ST131 from Australia. *Microb Genom* 2021;7:12.
  57. Petty NK, Ben Zakour NL, Stanton-Cook M, Skippington E, Totsika M, et al. Global dissemination of a multidrug resistant *Escherichia coli* clone. *Proc Natl Acad Sci* 2014;111:5694–5699.
  58. Toulouse JL, Edens TJ, Alejaldre L, Manges AR, Pelletier JN. Integron-associated DfrB4, a previously uncharacterized member of the trimethoprim-resistant dihydrofolate reductase B family, is a clinically identified emergent source of antibiotic resistance. *Antimicrob Agents Chemother* 2017;61:e02665–16.
  59. Shaw KJ, Rather PN, Hare RS, Miller GH. Molecular genetics of aminoglycoside resistance genes and familial relationships of the aminoglycoside-modifying enzymes. *Microbiol Rev* 1993;57:138–163.
  60. Lacotte Y, Ploy MC, Raheison S. Class 1 integrons are low-cost structures in *Escherichia coli*. *ISME J* 2017;11:1535–1544.
  61. Spurbeck RR, Dinh PC, Walk ST, Stapleton AE, Hooton TM, et al. *Escherichia coli* isolates that carry *vat*, *yfuA*, *chuA*, and *yfcV* efficiently colonize the urinary tract. *Infect Immun* 2012;80:4115–4122.
  62. Dale AP, Woodford N. Extra-intestinal pathogenic *Escherichia coli* (ExPEC): Disease, carriage and clones. *J Infect* 2015;71:615–626.
  63. Wright KJ, Hultgren SJ. Sticky fibers and uropathogenesis: bacterial adhesins in the urinary tract. *Future Microbiol* 2006;1:75–87.
  64. Ostolaza H, Goñi FM. Interaction of the bacterial protein toxin alpha-haemolysin with model membranes: protein binding does not always lead to lytic activity. *FEBS Lett* 1995;371:303–306.
  65. Wiles TJ, Kulesus RR, Mulvey MA. Origins and virulence mechanisms of uropathogenic *Escherichia coli*. *Exp Mol Pathol* 2008;85:11–19.
  66. Rodríguez I, Figueiredo AS, Sousa M, Aracil-Gisbert S, Fernández-de-Bobadilla MD, et al. A 21-year survey of *Escherichia coli* from Bloodstream Infections (BSI). *mSphere* 2021;6:e00868–21.
  67. Gibreel TM, Dodgson AR, Cheesbrough J, Fox AJ, Bolton FJ, et al. Population structure, virulence potential and antibiotic susceptibility of uropathogenic *Escherichia coli* from Northwest England. *J Antimicrob Chemother* 2012;67:346–356.
  68. Yamaji R, Rubin J, Thys E, Friedman CR, Riley LW, et al. Persistent pandemic lineages of uropathogenic *Escherichia coli* in a college community from 1999 to 2017. *J Clin Microbiol* 2018;56:e01834–17.
  69. Fibke CD, Croxen MA, Geum HM, Glass M, Wong E, et al. Genomic epidemiology of major extraintestinal pathogenic *Escherichia coli* lineages causing urinary tract infections in young women across Canada. *Open Forum Infect Dis* 2019;6:fz431.
  70. Doumith M, Day M, Ciesielczuk H, Hope R, Underwood A, et al. Rapid identification of major *Escherichia coli* sequence types causing urinary tract and bloodstream infections. *J Clin Microbiol* 2015;53:160–166.
  71. Dale AP, Woodford N. Extra-intestinal pathogenic *Escherichia coli* (ExPEC): Disease, carriage and clones. *J Infect* 2015;71:615–626.
  72. Bengtsson S, Naseer U, Sundsfjord A, Kahlmeter G, Sundqvist M. Sequence types and plasmid carriage of uropathogenic *Escherichia coli* devoid of phenotypically detectable resistance. *J Antimicrob Chemother* 2012;67:69–73.
  73. Sunde M, Simonsen GS, Slettemeås JS, Böckerman I, Norström M. Integron, plasmid and host strain characteristics of *Escherichia coli* from humans and food included in the norwegian antimicrobial resistance monitoring programs. *PLoS One* 2015;10:e0128797.
  74. Moura A, Soares M, Pereira C, Leitão N, Henriques I, et al. INTEGRALL: a database and search engine for integrons, integrases and gene cassettes. *Bioinformatics* 2009;25:1096–1098.
  75. Pitout JDD, DeVinney R. *Escherichia coli* ST131: a multidrug-resistant clone primed for global domination. *F1000Res* 2017;6:F1000 Faculty Rev-195.
  76. Johnson JR, Tchesnokova V, Johnston B, Clabots C, Roberts PL, et al. Abrupt emergence of a single dominant multidrug-resistant strain of *Escherichia coli*. *J Infect Dis* 2013;207:919–928.
  77. Gillings MR. Integrons: past, present, and future. *Microbiol Mol Biol Rev* 2014;78:257–277.
  78. Nemergut DR, Robeson MS, Kysela RF, Martin AP, Schmidt SK, et al. Insights and inferences about integron evolution from genomic data. *BMC Genomics* 2008;9:261.
  79. Wyrsh ER, Nesporova K, Tarabai H, Jamborova I, Bitar I, et al. Urban wildlife crisis: Australian silver gull is a bystander host to widespread clinical antibiotic resistance. *mSystems* 2022;7:e0015822.
  80. Clark AW, Durkin MJ, Olsen MA, Keller M, Ma Y, et al. Rural–urban differences in antibiotic prescribing for uncomplicated urinary tract infection. *Infect Control Hosp Epidemiol* 2021;42:1437–1444.
  81. Yau JW, Thor SM, Tsai D, Speare T, Rissel C. Antimicrobial stewardship in rural and remote primary health care: a narrative review. *Antimicrob Resist Infect Control* 2021;10.
  82. Gündoğdu A, Long YB, Vollmerhausen TL, Katouli M. Antimicrobial resistance and distribution of *intI* genes and integron-associated *intI* genes among uropathogenic *Escherichia coli* in Queensland, Australia. *J Med Microbiol* 2011;60:1633–1642.
  83. McInnes RS, Uz-Zaman MH, Alam IT, Ho SFS, Moran RA, et al. Metagenome-wide analysis of rural and urban surface waters

- and sediments in Bangladesh identifies human waste as a driver of antibiotic resistance. *mSystems* 2021;6:e0013721.
84. Ludden C, Coll F, Gouliouris T, Restif O, Blane B, et al. Defining nosocomial transmission of *Escherichia coli* and antimicrobial resistance genes: a genomic surveillance study. *Lancet Microbe* 2021;2:e472–e480.
  85. Mills EG, Martin MJ, Luo TL, Ong AC, Maybank R, et al. A one-year genomic investigation of *Escherichia coli* epidemiology and nosocomial spread at a large US healthcare network. *Genome Med* 2022;14:147.
  86. de Souza da-Silva AP, de Sousa VS, Martins N, da Silva Dias RC, Bonelli RR, et al. *Escherichia coli* sequence type 73 as a cause of community acquired urinary tract infection in men and women in Rio de Janeiro, Brazil. *Diagn Microbiol Infect Dis* 2017;88:69–74.
  87. Manges AR, Harel J, Masson L, Edens TJ, Portt A, et al. Multilocus sequence typing and virulence gene profiles associated with *Escherichia coli* from human and animal sources. *Foodborne Pathog Dis* 2015;12:302–310.
  88. Bourne JA, Chong WL, Gordon DM. Genetic structure, antimicrobial resistance and frequency of human associated *Escherichia coli* sequence types among faecal isolates from healthy dogs and cats living in Canberra, Australia. *PLoS One* 2019;14:e0212867.
  89. Liu X, Thungrat K, Boothe DM, Zhang Q. Multilocus sequence typing and virulence profiles in uropathogenic *Escherichia coli* isolated from cats in the United States. *PLoS One* 2015;10:e0143335.
  90. Valat C, Drapeau A, Beurlet S, Bachy V, Boulouis H-J, et al. Pathogenic *Escherichia coli* in dogs reveals the predominance of ST372 and the human-associated ST73 extra-intestinal lineages. *Front Microbiol* 2020;11:580.
  91. Melendez D, Roberts MC, Greninger AL, Weissman S, No D, et al. Whole-genome analysis of extraintestinal pathogenic *Escherichia coli* (ExPEC) MDR ST73 and ST127 isolated from endangered southern resident killer whales (*Orcinus orca*). *J Antimicrob Chemother* 2019;74:2176–2180.
  92. Becker S, Salden A, van Vliet AHM, Stegger M, Johannesen TB, Semmler T, et al. Genomic analysis of the zoonotic ST73 lineage containing avian and human extraintestinal pathogenic *Escherichia coli* (ExPEC). *Vet Microbiol* 2022;267:109372.
  93. Alhashash F, Wang X, Paszkiewicz K, Diggle M, Zong Z, et al. Increase in bacteraemia cases in the East Midlands region of the UK due to MDR *Escherichia coli* ST73: high levels of genomic and plasmid diversity in causative isolates. *J Antimicrob Chemother* 2016;71:339–343.
  94. Darling AE, McKinnon J, Worden P, Santos J, Charles IG, et al. A draft genome of *Escherichia coli* sequence type 127 strain 2009-46. *Gut Pathog* 2014;6:32.
  95. Nowak K, Fahr J, Weber N, Lübke-Becker A, Semmler T, et al. Highly diverse and antimicrobial susceptible *Escherichia coli* display a naïve bacterial population in fruit bats from the Republic of Congo. *PLoS One* 2017;12:e0178146.
  96. Jørgensen SL, Stegger M, Kudirkienė E, Lilje B, Poulsen LL, et al. Diversity and population overlap between avian and human *Escherichia coli* belonging to sequence type 95. *mSphere* 2019;4:e00333–18.
  97. Johnson TJ, Kariyawasam S, Wannemuehler Y, Mangiamela P, Johnson SJ, et al. The genome sequence of avian pathogenic *Escherichia coli* strain O1:K1:H7 shares strong similarities with human extraintestinal pathogenic *E. coli* genomes. *J Bacteriol* 2007;189:3228–3236.
  98. Moulin-Schouleur M, Répérant M, Laurent S, Brée A, Mignon-Grasteau S, et al. Extraintestinal pathogenic *Escherichia coli* strains of avian and human origin: link between phylogenetic relationships and common virulence patterns. *J Clin Microbiol* 2007;45:3366–3376.
  99. Mora A, López C, Dabhi G, Blanco M, Blanco JE, et al. Extraintestinal pathogenic *Escherichia coli* O1:K1:H7/NM from human and avian origin: detection of clonal groups B2 ST95 and D ST59 with different host distribution. *BMC Microbiol* 2009;9:132.
  100. Stephens CM, Skerker JM, Sekhon MS, Arkin AP, Riley LW. Complete genome sequences of four *Escherichia coli* ST95 isolates from bloodstream infections. *Genome Announc* 2015;3:e01241–15.
  101. Kudinha T, Johnson JR, Andrew SD, Kong F, Anderson P, et al. Distribution of phylogenetic groups, sequence type ST131, and virulence-associated traits among *Escherichia coli* isolates from men with pyelonephritis or cystitis and healthy controls. *Clin Microbiol Infect* 2013;19:E173–80.
  102. Sora VM, Meroni G, Martino PA, Soggiu A, Bonizzi L, et al. Extraintestinal pathogenic *Escherichia coli*: virulence factors and antibiotic resistance. *Pathogens* 2021;10:1355.
  103. Massot M, Daubié A-S, Clermont O, Jauréguy F, Couffignal C, et al. Phylogenetic, virulence and antibiotic resistance characteristics of commensal strain populations of *Escherichia coli* from community subjects in the Paris area in 2010 and evolution over 30 years. *Microbiology* 2016;162:642–650.
  104. Marin J, Clermont O, Royer G, Mercier-Darty M, Decousser JW, et al. The population genomics of increased virulence and antibiotic resistance in human commensal *Escherichia coli* over 30 years in France. *Appl Environ Microbiol* 2022;88:e0066422.
  105. Bonten M, Johnson JR, van den Biggelaar AHJ, Georgalis L, Geurtsen J, et al. Epidemiology of *Escherichia coli* bacteremia: a systematic literature review. *Clin Infect Dis* 2021;72:1211–1219.
  106. Messerer M, Fischer W, Schubert S. Investigation of horizontal gene transfer of pathogenicity islands in *Escherichia coli* using next-generation sequencing. *PLoS One* 2017;12:e0179880.
  107. Tyson GH, McDermott PF, Li C, Chen Y, Tadesse DA, et al. WGS accurately predicts antimicrobial resistance in *Escherichia coli*. *J Antimicrob Chemother* 2015;70:2763–2769.

### Five reasons to publish your next article with a Microbiology Society journal

1. When you submit to our journals, you are supporting Society activities for your community.
2. Experience a fair, transparent process and critical, constructive review.
3. If you are at a Publish and Read institution, you'll enjoy the benefits of Open Access across our journal portfolio.
4. Author feedback says our Editors are 'thorough and fair' and 'patient and caring'.
5. Increase your reach and impact and share your research more widely.

Find out more and submit your article at [microbiologyresearch.org](https://microbiologyresearch.org).