Current Research in Microbial Sciences 3 (2022) 100106

Contents lists available at ScienceDirect

Current Research in Microbial Sciences

journal homepage: www.sciencedirect.com/journal/current-research-in-microbial-sciences

Close genetic linkage between human and companion animal extraintestinal pathogenic *Escherichia coli* ST127

Paarthiphan Elankumaran [a], Glenn F. Browning [b], Marc S. Marenda [b], Cameron J. Reid [a], Steven P. Djordjevic [a,*]

[a] *iThree Institute, School of Life Sciences, Faculty of Science, University of Technology Sydney, Ultimo, NSW, Australia*
[b] *Asia-Pacific Centre for Animal Health, Department of Veterinary Biosciences, Melbourne Veterinary School, Faculty of Veterinary and Agricultural Sciences, University of Melbourne, Parkville and Werribee, Victoria, Australia*

ARTICLE INFO

Key words:
E. coli
ST127
Extraintestinal pathogenic *E. coli*
ExPEC
Genomic epidemiology
Phylogenomics
Companion animals
E. coli virulence
Antibiotic resistance
Interspecies *E. coli* transfer
One Health

ABSTRACT

Escherichia coli ST127, a recently emerged global pathogen noted for high virulence gene carriage, is a leading cause of urinary tract and blood stream infections. ST127 is frequently isolated from humans and companion animals; however, it is unclear if they are distinct or related populations of ST127. We performed a phylogenomic analysis of 299 *E. coli* ST127 of diverse epidemiological origin to characterize their population structure, genetic determinants of virulence, antimicrobial resistance, and repertoire of mobile genetic elements with a focus on plasmids. The core gene phylogeny was divided into 13 clusters, the largest of which (BAP4) contained the majority of human and companion animal origin isolates. This dominant cluster displayed genetic differences to the remainder of the phylogeny, most notably alternative gene alleles encoding important virulence factors including lipid A, flagella, and K capsule. Furthermore, numerous close genetic linkages (<30 SNPs) between human and companion animal isolates were observed within the cluster. Carriage of antimicrobial resistance genes in the collection was limited, but virulence gene carriage was extensive. We found evidence of pUTI89-like virulence plasmid carriage in over a third of isolates, localised to four of the major phylogenetic clusters. Our study supports global scale repetitive transfer of *E. coli* ST127 lineages between humans and companion animals, particularly within the dominant BAP4 cluster.

1. Introduction

E. coli is both a ubiquitous commensal species and important pathogen of vertebrates. There are more than 11,000 *E. coli* sequence types, but only a handful of pandemic lineages, including ST131, ST69, ST73, ST127, ST95 and ST117, are recognized human extraintestinal pathogenic *E. coli* (ExPEC) (Manges et al. 2019). Though genetic backgrounds such as phylogroups are broadly predictive of pathogenicity, a comprehensive understanding of genomic, host and environmental factors that facilitate expansion of these lineages from relative obscurity to globally disseminated pathogens is lacking. As most extraintestinal infections are caused by ExPEC resident among the host fecal commensal microbiota, holding a competitive advantage in commensal niches is one factor likely to influence the emergence of pandemic ExPEC. Mobile genetic elements (MGEs), particularly plasmids, that carry genes conferring virulence, antimicrobial resistance and expansion of metabolic capacity represent important genomic factors in this regard as they may simultaneously facilitate both commensal fitness and extraintestinal pathogenicity (Johnson 2021). One health genomic epidemiological approaches are required to account for the multiplicity of interactions between *E. coli* genomic backgrounds, MGEs, and their distribution across hosts and geographies (Soysal et al. 2016). These methods and the information they generate are key building blocks of future genomic surveillance systems that may allow tracking and prediction of emerging pathogens with relevance to human and animal health.

E. coli ST127 is a recently emerged pathogen noted for high virulence gene carriage (Gibreel et al. 2012; Darling et al. 2014). ST127 is a leading cause of human urinary tract (Banerjee et al. 2013; Alghoribi et al. 2015; Ciesielczuk et al. 2016; Yamaji et al. 2018; Manges et al. 2019) and blood stream infections (Hastak et al. 2020) in most regions of the world and has been: i) associated with necrotizing enterocolitis in

* Corresponding author.
E-mail addresses: Paarthiphan.Elankumaran@student.uts.edu.au (P. Elankumaran), Glenfb@unimelb.edu.au (G.F. Browning), Mmarenda@unimelb.edu.au (M.S. Marenda), Cameron.Reid@uts.edu.au (C.J. Reid), Steven.Djordjevic@uts.edu.au (S.P. Djordjevic).

https://doi.org/10.1016/j.crmicr.2022.100106

Available online 23 January 2022

pre-term infants (Ward et al. 2016); ii) cultured from neonatal nasogastric feeding tubes (Alkeskas et al. 2015); and iii) transmitted sexually between males with febrile UTI and their female partners (Ulleryd et al. 2015). In returning travelers, multiple drug resistant variants of ST127 can colonize the gut of healthy humans and be shed for considerable periods in the absence of antibiotic selection (Zurfluh et al. 2016). Isolation from companion animals (Johnson et al. 2008; Bourne et al. 2019; Kidsley et al. 2020a; Kidsley et al. 2020b), whales (Melendez et al. 2019), bats from the Congo (Nowak et al. 2017), and river water in Japan (Gomi et al. 2017) indicate it is adept at colonizing diverse hosts. ST127 are also known to be particularly lethal in a *Galleria mellonella* model of infection (Alghoribi et al. 2014). These attributes rank ST127 as a major threat to the health of humans and companion animals.

The influence of mobile genetic elements, particularly plasmids, on the evolution of pathogenic *E. coli* is increasingly acknowledged. Plasmids belonging to the F replicon family are conspicuous in this regard. F plasmids are a heterogenous group of typically conjugative plasmids that often carry virulence-associated genes (VAGs), antimicrobial resistance genes (ARGs) and insertion sequences (IS) that may facilitate rapid evolution via gene gain or loss (Johnson 2021). Two major groups of F plasmids, commonly found in ExPEC are ColV plasmids and ColIa/pUTI89-like plasmids. Both groups carry operons that function in iron-acquisition and are associated with both intestinal fitness and extraintestinal virulence (Cusumano et al. 2010; Johnson 2021). Whilst ColV plasmids frequently carry large clusters of ARGs (McKinnon et al., 2018; Moran and Hall 2018; Cummins et al., 2022), pUTI89-like plasmids rarely carry such genes and are also negatively associated with strain carriage of ARGs (Stephens et al. 2017; Cummins et al. 2022). ColV plasmids are highly associated with poultry source *E. coli,* such as ST117 (Cummins et al. 2019; McKinnon et al. 2018; Cummins et al. 2022) but are also typical of broad host range *E. coli,* such as specific lineages of ST131-*H*22 (Liu et al. 2018; Reid et al. 2019). By contrast, *E. coli* that carry pUTI89-like plasmids appear to be mostly restricted to human hosts via dominant STs such as ST95 and ST131 clades A and B (Cummins et al. 2022; Stephens et al. 2017; Li et al. 2021). The presence and/or role of F plasmids and their respective subtypes in ST127 however, is yet to be elucidated.

Snapshot whole genome sequencing (WGS) studies of *E. coli* isolated from domestic animals with urinary tract disease identified ST127 (O6:H31) from Australian cats (Kidsley et al. 2020b) and dogs (Kidsley et al. 2020a), however a larger study of ST127 that includes isolates from diverse hosts and sources is lacking. ST127 *E. coli* featured prominently (22/399; 5.51%) in a large WGS study of 377 Australian *E. coli* isolates from dogs with urinary tract disease (our unpublished data). Here we performed a phylogenomic analysis of a global collection of 299 *E. coli* ST127 isolated from different hosts and geographical regions to characterize population structure, source-lineage linkages and distinguishing genomic features of this important pathogenic lineage.

2. Materials and methods

2.1. Isolates used in this study

The total collection under analysis comprised 299 genome sequences of ST127 *E. coli.* Twenty-two originated from a collection of *E. coli* isolated from diseased canine hosts, seven from collections we previously published and the remaining 270 from Enterobase.

The 22 genome sequences described for the first time here were from a collection of 377 *E. coli* isolated from dogs presenting with various extraintestinal pathologies, obtained from the Melbourne Veterinary School, University of Melbourne, Australia. The isolates were sequenced at the University of Technology sequencing facility as described below. These isolates carry a "MVC" (meaning "Melbourne Veterinary Collection") prefix followed by a 1 to 3-digit numeral specifying individual isolates from the collection. Seven further sequences originated from human extraintestinal infections previously published by our group and can be accessed via accessions in Table S1.

270 sequences were obtained from Enterobase (http://enterobase.warwick.ac.uk) accessed on 30th April, 2020 (Zhou et al. 2020). Initially the database was queried for *E. coli* belonging to ST127 and the corresponding accession numbers and relevant metadata were downloaded. These data were filtered to exclude isolates without sound accession numbers, source details, year of isolation, country, and continent of origin. This final list of filtered samples was used to query the National Center for Biotechnology Information (NCBI) and European Bioinformatics Institute (EBI) sequence read databases. Short read sets for all isolates were downloaded with parallel-fastq-dump (https://github.com/rvalieris/parallel-fastq-dump). These 270 genome sequences were named using their NCBI or EBI accession numbers (Table S1).

2.2. Genomic DNA isolation, whole genome sequencing and assembly

E. coli ST127 isolates from the Melbourne Veterinary Collection were freshly cultured onto LB agar plates and a single colony used to inoculate 5 ml of sterile LB medium. Following overnight culture, total cellular DNA was extracted using the ISOLATE II Genomic DNA (Bioline) kit following the manufacturer's standard protocol for bacterial cells and stored by refrigerating at 4 °C. Library preparation was done by the iThree Core Sequencing Facility, University of Technology Sydney, following the adapted Nextera Flex library preparation kit process, Hackflex (Gaio et al. 2019). Briefly, genomic DNA was quantitatively assessed using the Quant-iT picogreen dsDNA assay kit (Invitrogen, USA). Each sample was normalised to a concentration of 1 ng/μl. A 10 ng sample of DNA was used for library preparation. After tagmentation, DNA was amplified using the facility's custom designed i7 and i5 barcodes, with 12 cycles of PCR. Due to the number of samples, the quality control for the samples was done by sequencing a pool of samples using the MiSeq V2 nano kit – 300 cycles. Briefly, after library amplification, 3 μl of each library was pooled into a library pool. The pool was then cleaned up using SPRIselect beads (Beckman Coulter, USA) following the Hackflex protocol. The pool was sequenced using the MiSeq V2 nano kit (Illumina, USA). Based on the sequencing data generated, the read count for each sample was used to identify the failed libraries (i.e. libraries with less than 100 reads), and normalised to ensure equal representation in the final pool. The final pool was sequenced on one lane of an Illumina Novaseq S4 flow cell, 2 × 150 bp at Novogene (Singapore). The quality of reads generated were confirmed with fastp (0.20.1).

2.3. Genome assembly and gene screening

A modular analysis pipeline known as pipelord2, implemented with the Snakemake workflow management system was used to perform primary bioinformatic analysis (Koster and Rahmann 2012). This pipeline is freely available to download from https://github.com/maxlcummins/pipelord2_0. Default settings are used unless otherwise stated. Firstly, Kraken2 was applied to the sequence reads to confirm all genomes were *E. coli.* Draft genomes were then assembled with Shovill 1.0.4 (https://github.com/tseemann/shovill), with default settings and assembly-stats run to confirm the quality of the assemblies (https://github.com/sanger-pathogens/assembly-stats). Assemblies with >800 contigs or total length <4.5Mbp or >6.5Mbp were excluded. MLST 2.19.0 (https://github.com/tseemann/mlst) was used to confirm all genomes belonged to ST127 (Jolley and Maiden 2010). ABRicate 1.0.1 (https://github.com/tseemann/abricate) was used to screen draft genomes for genes from several publicly available and custom in-house databases. Public databases used were CARD, VFDB, PlasmidFinder, SerotypeFinder and ISFinder (Siguier et al. 2006; Carattoli et al. 2014; Chen et al. 2016; Ingle et al. 2016; Jia et al. 2017). The custom database included the set of genes used to infer ColV plasmid carriage (see below) and additional virulence genes. This is available at https://github.com/maxlcummins/custom_DBs. ABRicate was also used to align assemblies to

the reference pUTI89 plasmid from the E. coli strain UTI89, sourced from GenBank (gb | NC_007941). pMLST was performed with the pMLST tool available at https://bitbucket.org/genomicepidemiology/cge-tools-docker/src/master/ (Carattoli and Hasman 2020). AMR-associated SNPs were identified with PointFinder (Zankari et al. 2017). Finally, gene screening results are summarized by abricateR (https://github.com/maxlcummins/abricateR) with a gene being considered present at 95% length and 90% nucleotide identity.

2.4. Criteria for inference of plasmid presence

The presence of a ColV type plasmid was inferred using criteria previously described by Liu *et al,* 2018 (Liu et al. 2018). The presence of a pUTI89-like plasmid was inferred if a given assembly mapped to ≥90% of the pUTI89 sequence at ≥90% identity or if the isolate was determined by pMLST to carry the F29:A-:B10 RST combination, which is characteristic of pUTI89-like plasmids.

2.5. Phylogenetic and SNP distance analyses

The assembled *E. coli* ST127 genomes and the genome of outgroup strain MVC107 (ST372) were annotated using prokka 1.14.6 (Seemann 2014). The core and pangenome was then determined with Roary 3.13.0 with default settings and paralog splitting on (Page et al. 2015). The resulting core gene alignment of 3,266,764 bp was then used as the basis for subsequent analyses. IQTree 2.0.3 was used to infer a maximum-likelihood phylogenetic tree using the GTR+F+R substitution model and 1000 bootstrap replicates (Nguyen et al. 2015). FigTree 1.4.4 (https://github.com/rambaut/figtree) was used to root the tree on the outgroup sequence, and subsequently remove it for tree visualization. snp-sites 2.5.1 was run on the core gene alignment to identify core variable SNP sites, resulting in a core SNP alignment of 30,896 bp (Page et al. 2016). Pairwise SNPs were extracted from the core SNP alignment with snp-dists 0.6.3 (https://github.com/tseemann/snp-dists). Fastbaps was used with a 'baps' prior to define clusters of isolates based on the core gene alignment and maximum-likelihood tree (Tonkin-Hill et al. 2019).

2.6. Genome wide association studies (GWAS)

Scoary 1.6.16 was used to determine associations between fastbaps cluster membership and genes in the ST127 pangenome (Brynildsrud et al. 2016). A Benjamini-Hochberg-adjusted p-value cutoff of 1E-30 was used to determine significant associations. Biological process terms associated with the identified genes were derived from UniProt entries for each gene.

2.7. Data analysis and visualization

A custom R script was written in RStudio 1.4.1106 with R 4.0.5 to perform secondary analysis on the data generated by pipelord2 and via the phylogenetic methods, and to generate publication figures. The sequence of plasmid pUTI89 was visualised with SnapGene® Viewer (Version 5.0.7, GSL Biotech LLC). Microsoft PowerPoint was used to compile elements of Figs. 2; Figure 4 and Figure 5. The data analysis and visualization script is available at https://github.com/CJREID/ST127 and can be used to reproduce all secondary analysis. R package versions used therein are available within the README.md document in the code repository.

2.8. Genomic data deposition

Melbourne Veterinary Collection (MVC) genomes were deposited in GenBank and the Sequence Read Archive under the BioProject PRJNA678027. Orange Base Hospital (HOS) genomes were deposited in GenBank under the BioProject PRJNA623470. Sydney Adventist Hospital genomes were deposited in GenBank under the BioProject PRJNA732725. Individual accession numbers can be found in Table S1.

3. Results

3.1. The study collection

The study collection consisted of 299 individual ST127 *E. coli* isolates of diverse epidemiological origins. The isolates were sourced from humans, companion animals, livestock, wild animals, aquatic organisms, abiotic environments, and food. Five continents and 18 countries were represented, with a temporal distribution of 1977–2019 (Fig. 1 and Table S1). Human and companion animals were dominant sources (164/299, 54.8% and 85/299, 28.43% respectively), though companion animal isolates only originated from Australia, Canada and United States, whilst human isolates originated from 17/18 countries represented (Table S1).

3.2. Phylogenetic relationships between the ST127 isolates

The core and pangenome sizes of the study collection were determined by Roary with the 299 *E. coli* ST127 prokka-annotated draft genomes and outgroup strain MVC107 (ST372). The full pan-genome consisted of 20,349 genes. The core genome (present in ≥ 99% of genomes) consisted of 3467 genes, leaving 16,882 genes within the accessory genome. Note that while the ST372 outgroup strain will slightly increase the size of the accessory genome, it will not affect the estimation of the core due to the 99% gene presence threshold used to define the core.

A maximum likelihood phylogeny was inferred with IQTree from a multiple alignment of the core genes identified by Roary rooted on *E. coli* ST372 strain MVC107 as an outgroup (Fig. 2; MVC107 tip removed). The phylogeny was divided into 13 clusters by fastbaps analysis designated BAP1–13. The largest cluster was BAP4, which contained more than half of all sequences (163/299, 54.51%) and was followed by BAP7 (34, 11.37%), BAP6 (28, 9.36%), BAP10 (24, 8.03%), BAP3 (14, 4.68%) and BAP1 (12, 4.01%) (Fig. 3a). The remaining clusters contained less than 10 sequences each. BAP4 contained more than half of human and companion animal sourced sequences (101/164, 61.60% and 46/85, 54.12%, respectively), though human and companion animal sequences were also present together in seven other clusters (Fig. 3a). Five continents and 14 countries were represented within BAP4 (Table S1). pUTI89-like plasmids were identified in BAP4, BAP6, BAP7 and BAP8 in variable proportions and were also present in all sources except aquatic (Fig. 3b-c; see *3.4.3. Plasmids* below).

3.3. SNP distances between the ST127 isolates

To identify cases of closely related isolates from epidemiologically unrelated sources, we calculated pairwise SNP distances between all isolates and filtered pairs differing by ≤30 SNPs. This analysis identified 57 unique isolate pairs, 26 of which were between companion animal and human isolates (Fig. 4). Sixteen companion animal:wild animal pairs and seven human:wild animal pairs were also identified. Most pairs occurred within the dominant BAP4 cluster (39/57). Australian companion animal isolates were linked with isolates from the United Kingdom, United States, Canada, Denmark, Sri Lanka and Oman.

3.4. Genetic features of ST127

We used ABRicate to screen ST127 genomes for antimicrobial resistance genes (ARGs), virulence-associated genes (VAGs), plasmid replicons and insertion sequences. These results were summarised as heatmaps mapped to the core gene phylogeny (Figs. S1–3, Table S1)

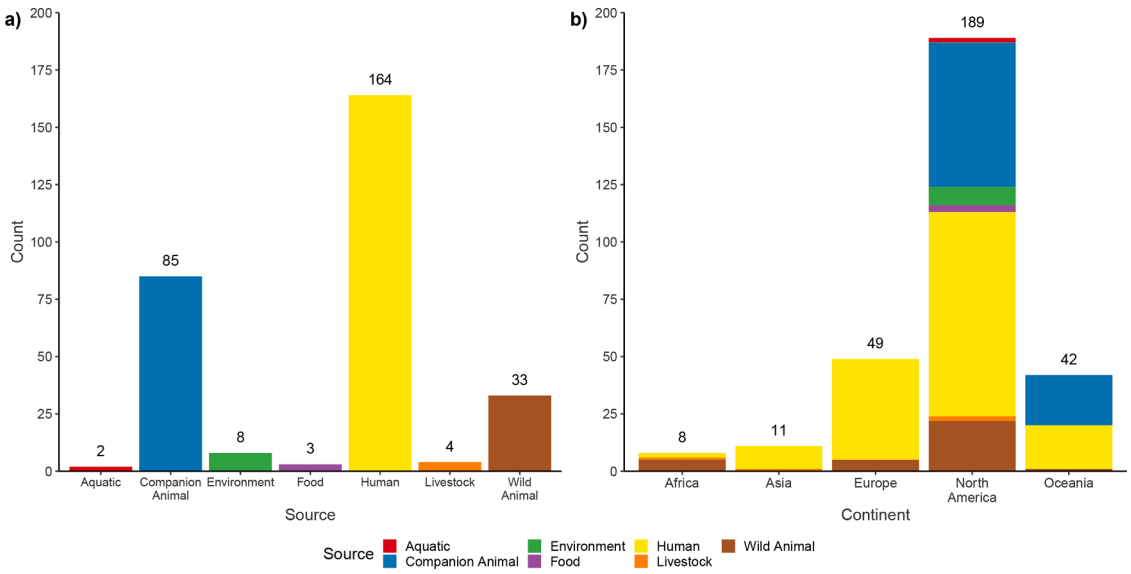

Fig. 1. Source and geographic distribution of ST127 isolates; a) Count of sequences per source and b) count of sequences by continent, stratified by source.

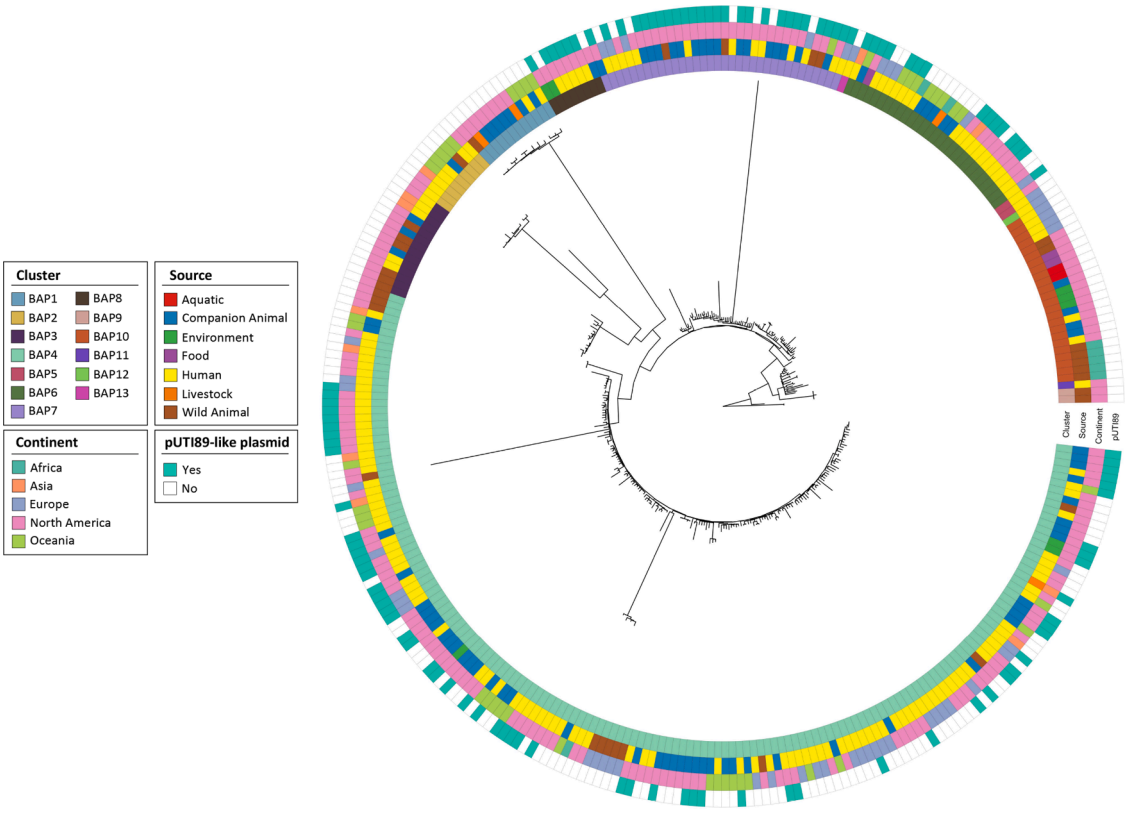

Fig. 2. Core gene based maximum-likelihood phylogenetic tree for the 299 *E. coli* ST127 isolates from the study collection. Coloured rings from inner to outermost display cluster defined by fastbaps, isolate source, continent of origin and inferred presence of pUTI89-like plasmids.

3.4.1. Antimicrobial resistance genes (ARGs)

A total of 58 distinct antimicrobial resistance genes (ARGs) were identified with no obvious linkage between any genes and defined clusters in the phylogeny. The number of resistance genes identified per isolate ranged from 0 to 11, with an average of 1.33 and median of 0. Only seven genes were carried at rates of greater than 5%; these included $bla_{\text{TEM-1B}}$ (64/299, 21.40%), *tet*(B) (47/299, 15.72%), *sul1* (36/299, 11.71%), *sul2* (28/299, 9.36%), *aph(3′′)-Ib/strA* (25/299, 8.36%), *aph-(6)-Id/strB* (24/299, 8.03%) and *catA1* (20, 6.69%). CTX-M type ESBL genes were rare yet diverse and included $bla_{\text{CTX-M-3}}$ (7/299; 2.34%), $bla_{\text{CTX-M-14}}$ (4/299, 1.34%), $bla_{\text{CTX-M-15}}$ (5/299, 1.67%) and $bla_{\text{CTX-M-55}}$ (1/299, 0.33%). In addition, $bla_{\text{CARB-2}}$ (2/299, 0.67%), $bla_{\text{OXA-1}}$ (3/299, 1%), $bla_{\text{OXA-48}}$ (2/299, 0.67%) and bla_{Z} (PC1 variant, 2/299, 0.67%) were identified, albeit rarely. The class 1 integron integrase gene *intI1* was present in 34 isolates (11.71%) and was found on the same scaffold as ARGs in all cases (Table S2). Eight *dfrA* trimethoprim resistance gene variants were present on 17 *intI1* positive (*intI1*+) scaffolds, whilst sulfonamide resistance gene *sul1*, a typical component

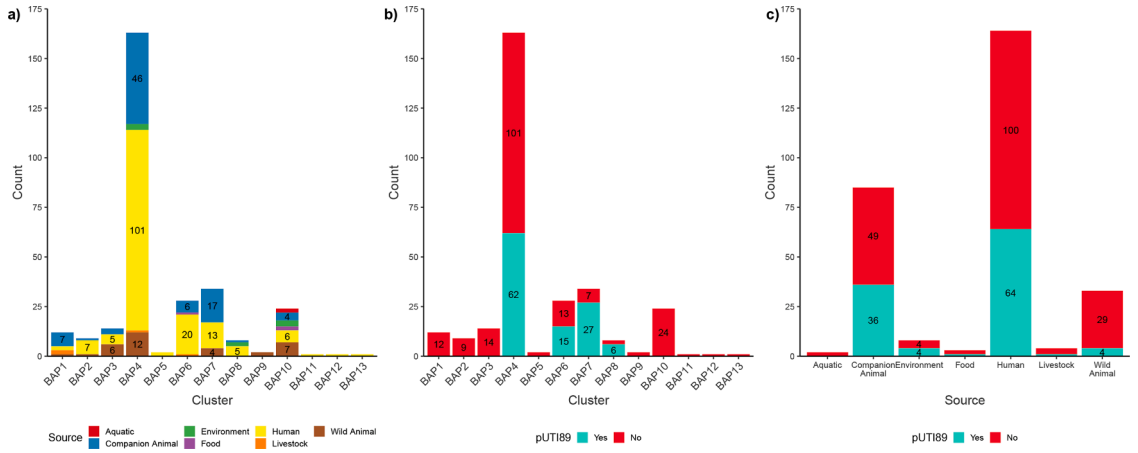

Fig. 3. Summary of clusters, sources and pUTI89-like plasmid carriage; a) count of sequences per cluster stratified by source, b) stratified by pUTI89-like plasmid carriage and c) count of sequences per source stratified by pUTI89-like plasmid carriage.

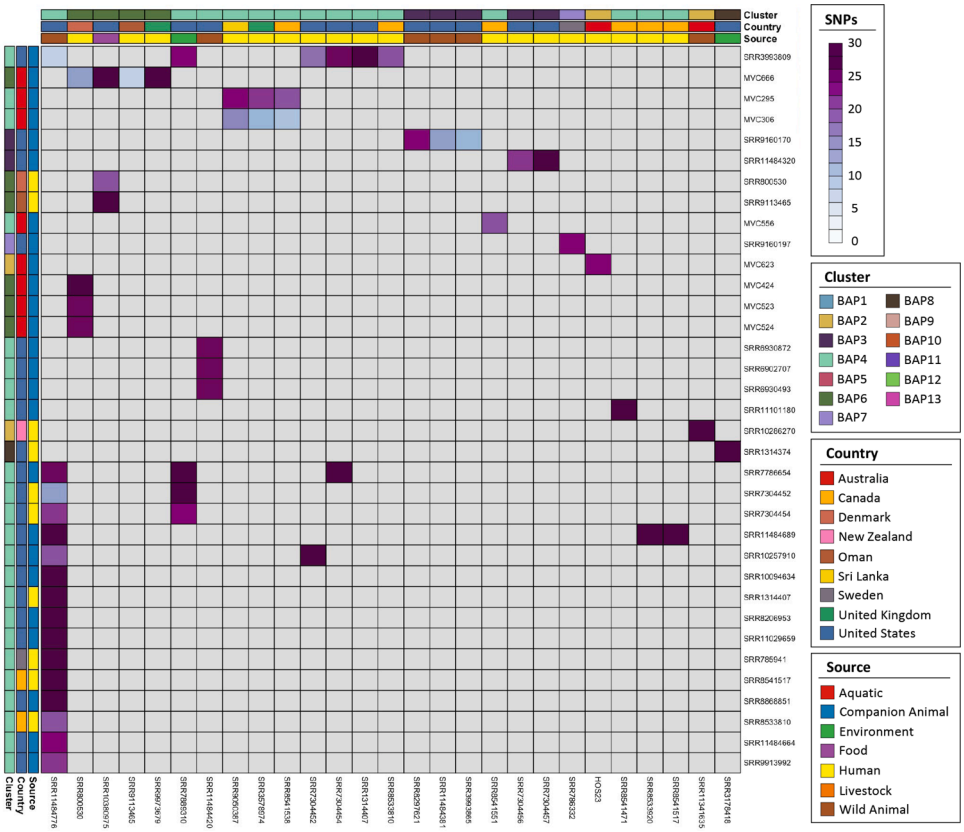

Fig. 4. Heatmap showing pairwise SNP distances ≤30 SNPs for human and companion animal isolates. Metadata for cluster, country and source is displayed as row and column annotations.

of the classical class 1 integron structure was present on 26 *intI1*+ scaffolds. Reflective of this co-localization data, isolates that were *intI1*+ carried a significantly higher average number of ARGs than *intI1*- isolates. Overall, our results indicate that a small subset of ST127 have acquired a variety of ARGs and integron cassette arrays, however these are not characteristic of its general evolution or the evolution of its sub-lineages.

3.4.2. *Virulence associated genes (VAGs)*

Virulence-associated gene profiles were extensive and relatively conserved across the phylogeny. We identified 163 genes from VFDB, and 61 of these were present in ≥99% of isolates. The number of VAGs per isolate ranged from 55 to 113, with an average of 94.06 and median 95. It should be noted that some of these genes are nearly ubiquitous in *E. coli* and their specific alleles, which were not characterised here, may be more relevant to actual virulence expression than simple gene presence or absence. Highly conserved virulence-associated gene loci (present in ≥95% of isolates) included enterobactin (*ent*), ferrienterobactin (*fep*), heme transport locus (*chu*), yersiniabactin (*ybt, fyuA, irp2*), K1 capsule (*kps*) and P fimbriae (*pap*). Other notable genes and loci included bacteriocin-like genotoxin *usp* (299/299, 100%), outer membrane protease *ompT* (297/299, 99.33%), *iss* conferring increased serum survival (288/299, 96.32%), serine protease *vat* (281/299, 93.98%) and salmochelin (*iroN;* 217/299, 72.58%). Interestingly, most genes of the *sfa*

operon encoding S fimbriae were present in more than 200 isolates, however major subunit *sfaA* was only present in 3 isolates. This virulence data underscores the virulence threat posed by ExPEC ST127 and is typical of a member of *E. coli* phylogroup B2.

3.4.3. Plasmids

Screening with the PlasmidFinder database indicated that FII and FIB replicon types were frequently observed in the collection. We therefore performed F plasmid replicon sequence typing (RST) to clarify these results with higher resolution. This analysis revealed 28 unique F RSTs in the collection with all but one present in less than 5% of isolates. The top three RSTs were F29:A-:B10 (103/299, 34.45%), F51:A-:B10 (14/299, 4.68%), and F2:A-:B- (9/299, 3.01%). The F29:A-:B10 RST is indicative of carriage of a pUTI89-like plasmid, which are abundant in various major lineages of uropathogenic *E. coli.* To complement this data we aligned ST127 sequences to the reference sequence of pUTI89 (gb | NC_007941) and considered a pUTI89-like plasmid to be present if an isolate carried the F29:A-:B10 allele combination and/or the sequence displayed a minimum nucleotide sequence identity of 90% over a minimum 90% of the reference sequence (Fig. 5). This analysis indicated that 104/299 (34.78%) isolates carried a pUTI89-like plasmid; 97 of which were F29:A-:B10, and seven others that were F2:A-:B10 (three), F29:A-:B- and F29:A8:B10 (two each). Six F29:A-:B10 isolates did not meet the 90% coverage threshold, displaying coverage ranging from 78–87%, however we have opted to consider these as pUTI89-like carriers in our analysis to allow for deletions in otherwise closely-related plasmids. The final pUTI89-like plasmid carriage was therefore considered to be 110/299 (36.79%).

Other replicons identified by PlasmidFinder included those indicative of small Col-type plasmids; *Col156* (152/299, 50.84%), *Col-MG828* (33/299, 11.04%), *Col-BS512* (11/299, 3.68%), and *ColpVC* (9/299, 3.01%). Other replicons included IncB/O/K/Z (22/299, 7.36%), IncI1 (12/299, 4.01%), IncY (8/299, 2.68%), IncI2 (6/299, 2.01%). All other replicons were present in less than 2% of isolates.

3.5. Associations between genetic features

It has been previously noted that *E. coli* that carry pUTI89-like plasmids are negatively associated with carriage of ARGs, whilst those that carry *intI1* are positively associated with carriage of ARGs. To test this in ST127, we compared the mean of total ARGs per isolate by pUTI89 and *intI1* status (Fig. 6). As expected, pUTI89+ isolates carried significantly less ARGs than pUTI89- isolates (mean 0.97 vs 1.51; Wilcoxon test; p=0.035), whilst *intI1*+ isolates carried significantly more ARGs than *intI1*- counterparts (mean = 4.72 vs 0.89; Wilcoxon test; p<2.2e-16). Despite the former result, 15 pUTI89+ strains carried between three and ten ARGs.

3.6. Genes associated with the dominant cluster of ST127

To examine genomic features associated with different evolutionary clusters of ST127, we carried out a GWAS analysis with Scoary using the BAP clusters as categorical traits. We utilized an adjusted p-value threshold of 1E-30, to identify only the most highly cluster-associated genes. Fifteen genes in total were identified and fourteen of these were associated with the largest cluster BAP4 (Table 1). Genes identified therein primarily encode proteins involved in cell surface functions (lipid A biosynthesis, flagellar biosynthesis, and K capsule) and carbohydrate metabolism (glyoxylate cycle, tricarboxylic acid cycle). Most genes identified here were present in a high proportion of ST127 strains and this analysis indicates that specific alleles of these genes are associated with BAP4 ST127.

4. Discussion

This phylogenomic study of *E. coli* ST127 globally has illuminated some key genomic characteristics of this important pathogenic lineage and provided avenues for further investigation.

Firstly, the dominance of geographically diverse human and companion animal isolates originating from cases of extraintestinal disease across the breadth of the phylogeny confirms that ST127 is an important pathogen regardless of sub-lineage (Gibreel et al. 2012; Fibke et al. 2019). At a sub-lineage level, the incidence of both geographically proximal and distal human and companion animal isolates that differ by less than 30 core SNPs strongly suggests repetitive large-scale transmission of ST127 between these hosts on a global scale. This is reflective of previous reports that document host sharing, persistence and multiple transmission events of ExPEC between companion animals and humans (Johnson et al. 2000; Johnson et al. 2008a; Johnson et al. 2008b; Bourne et al. 2019). Recent studies highlighting the frequency of isolation of *E. coli* ST127 from companion animals and from humans with extraintestinal infections, including bacteraemia (Gibreel et al. 2012; Horner et al. 2014; Riley 2014; LeCuyer et al. 2018; Flament-Simon et al. 2020; Kidsley et al. 2020a; Kidsley et al. 2020b), also support this contention and suggest *E. coli* ST127 is adept at colonizing and infecting both humans and companion animals. Only four isolates originated in livestock, suggesting food production is unlikely to have a major influence on the evolution and dissemination of ST127 as is thought to be the case for other prominent STs (Reid et al. 2019; Reid et al. 2020). By contrast, wild animal origin sequences did feature with some close genetic linkages to human and companion animal isolates, though their relevance is yet to be fully appreciated (Smalla et al. 2018; Nesporova

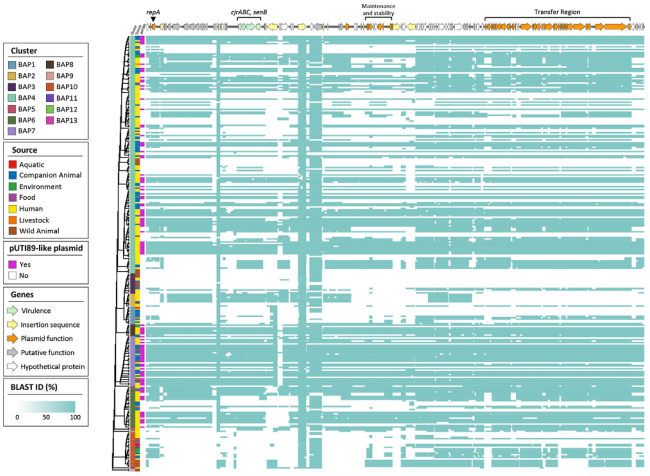

Fig. 5. Binned BLAST alignment of ST127 sequences to virulence plasmid pUTI89 aligned to phylogeny.

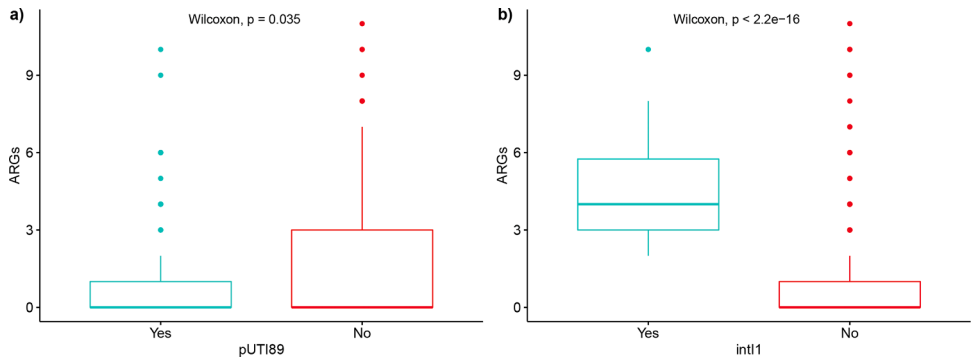

Fig. 6. Wilcoxon test comparison of mean ARG carriage by a) pUTI89 carriage and b) *intI1* carriage.

Table 1
Genes over-represented in the BAP4 cluster (n=163) of ST127. *Benjamini-Hochberg adjusted p-value for multiple comparison.

Gene	Annotation	Biological Process	BAP4 (n=163)	p-value*
lpxD	UDP-3-O-(3-hydroxymyristoyl) glucosamine N-acyltransferase	Lipid A biosynthesis	161	5.61E-81
flhB	Flagellar biosynthetic protein FlhB	Flagellar biosynthesis	160	1.63E-75
cheZ	Protein phosphatase CheZ	Flagellum-dependent swarming motility	160	1.74E-72
galE	UDP-glucose 4-epimerase	Carbohydrate metabolism	150	2.65E-65
kpsT	Polysialic acid transport ATP-binding protein KpsT	Polysialic acid transmembrane transport	162	2.16E-61
kpsM	Polysialic acid transport protein KpsM	Polysialic acid transmembrane transport	162	2.16E-61
glcA	Glycolate permease GlcA	Glycolate transmembrane transport	158	1.12E-60
glcB	Malate synthase G	Glyoxylate catabolic process	160	2.71E-60
group_5248	ISL3 family transposase ISEc53	Transposase	153	1.28E-52
ybhI	Inner membrane protein YbhI	Transmembrane transport	125	5.22E-41
dctD_1	C4-dicarboxylate transport transcriptional regulatory protein DctD	Transcription regulation	125	8.96E-40
sucB	Dihydrolipoyllysine-residue succinyltransferase component of 2-oxo-glutarate dehydrogenase complex	Tricarboxylic acid cycle	126	5.12E-38
dctD_2	C4-dicarboxylate transport transcriptional regulatory protein DctD	Transcription regulation	134	2.56E-33
sucC	Succinate–CoA ligase [ADP-forming] subunit beta	Tricarboxylic acid cycle	125	4.17E-33

et al. 2020; Wyrsch et al. 2020). Our observation that nearly all isolates in the collection were serotype O6:H31 and the fact that most of the *E. coli* isolates listed in Enterobase that have an O6:H31 serotype belong to ST127 indicates that a sizeable proportion of O6:H31 isolates historically identified in humans, dogs, cats and marine mammals were likely ST127 (Maluta et al. 2012; Melendez et al. 2019; Kidsley et al. 2020a; Kidsley et al. 2020b). Further, a previous study also found high similarity of virulence and PFGE profiles between *E. coli* ST127 isolates belonging to serotype O6:H31 from humans and dogs, indicating that ST127 successfully transfers between these species (Johnson et al. 2008).

Identification of a large cluster in the phylogeny (designated BAP4), which encompassed over half the collection is evidence for a globally dominant sub-lineage of ST127 that is worthy of further investigation. Sub-population structure among pandemic lineages of ExPEC is not without precedent; the example of ST131 being pertinent (Petty et al. 2014; Li et al. 2021; Bonnet et al. 2021). Understanding the factors responsible for emergence of successful STs and expansion of dominant sub-lineages within such STs is critical to understanding the evolution of commensal and pathogenic *E. coli* alike. To identify accessory genes over-represented in the dominant BAP4 cluster of ST127, we performed a GWAS-style analysis with Scoary. This analysis identified 14 over-represented genes, most of which encode proteins involved in two major biological processes; cell surface functions and carbohydrate metabolism. Among the cell surface functions, we identified gene alleles involved in Lipid A (*lpxD*) and K1 capsule biosynthesis (*kpsT, kpsM*), both of which have well known functions in *E. coli* virulence via immune evasion (Sarkar et al. 2014). Additionally, two genes involved in expression of flagellum (*flhB, cheZ*), a key motility factor were also identified (Lüthje and Brauner 2014). Regarding carbohydrate metabolism, genes involved in the glyoxylate (*glcA, glcB*) and tricarboxylic acid (TCA) cycles (*sucB, sucC*) featured. The glyoxylate cycle allows gluconeogenesis when only C_2 carbon compounds, such as ethanol and acetate, are available – bypassing decarboxylation steps of the TCA to produce malate and succinate (Kornberg 1966). Interestingly, *sucB* and *sucC* are involved in the TCA steps that are bypassed by the glyoxylate pathway. The over-representation of specific *glc* and *suc* gene alleles in the BAP4 may therefore be interrelated via repetitive exposure of the lineage to a limited variety of carbon sources. Broadly, these results suggest alterations to key virulence and metabolic traits that may provide BAP4 ST127 with an advantage over other sub-lineages of ST127 and explain its prominence.

Consistent with earlier reports of ST127, our cohort of ST127 genomes showed limited carriage of ARGs (Croxall et al. 2011; Banerjee et al. 2013; Hertz et al. 2016; Yamaji et al. 2018). However, this was not the case for ST127 that carried the *intI1* integrase gene. *intI1*+ isolates carried an average of 4.62 ARGs compared to 0.89 for *intI1*- isolates. This is consistent with the ability of class 1 integrons to capture and express multiple resistance genes and reside within large mosaic resistance regions (Reid et al. 2015); phenomena further supported by the co-occurrence of ARGs on assembly scaffolds that contained the *intI1* gene. Overall, this demonstrates that despite the low carriage of ARGs by ST127 in general, resistant strains might emerge via integron acquisition. Further work to determine transposons and plasmids responsible

for the carriage of integrons within ST127 would be beneficial as the evolution of AMR within a highly pathogenic lineage of *E. coli* is concerning.

In contrast to the ARG profiles, VAG profiles in our cohort was extensive and relatively conserved across the phylogeny. Several studies have found that ST127 carries a high virulence gene load and that infections caused by ST127 are often associated with more severe disease (Croxall et al. 2011; Gibreel et al. 2012; Beyrouthy et al. 2013; Alghoribi et al. 2014; Salipante et al. 2015). It is notable that, in a *Galleria mellonella* model of infection, *E. coli* ST127 is more virulent than other pandemic ExPEC lineages, including ST131, ST95, ST69 and ST73. Perhaps, with the exception of *E. coli* ST131, most of the dominant *E. coli* ExPEC lineages have risen to prominence without the necessity to acquire a large arsenal of antibiotic resistance genes (Kallonen et al. 2017; Yamaji et al. 2018; Hastak et al. 2020). A combination of factors, including carriage of extensive VAGs and intrinsic biological fitness (Yamaji et al. 2018), is likely to play a significant role in their global success. In this regard, *E. coli* ST127 is similar to *E. coli* ST73 and ST95, as they are well established extraintestinal pathogens with limited carriage of ARGs (Kallonen et al. 2017; Stephens et al. 2017; Bogema et al., 2020).

The role of F type plasmids in the evolution of AMR, virulence and fitness in ExPEC is increasingly apparent (Johnson 2021). Major F plasmid types involved in virulence include ColV plasmids and ColIa/pUTI89-like plasmids. Only a single isolate met the criteria for ColV carriage, though it curiously lacked an F replicon. By contrast, pUTI89-like plasmids were identified in over one third (110/299) of isolates, nested within four of the major evolutionary clusters. It is notable that one of the earliest pUTI89-like plasmids found to be associated with cystitis originated in an O6:H31 strain (DebRoy et al. 2010). Epidemiological data suggests that ColV and pUTI89-like plasmid types have different host distributions, in terms of both *E. coli* STs and eukaryotic host. ColV plasmids have a broad host range including humans, with major reservoirs in food animals, particularly poultry in association with ST117 and ST131-*H*22, whereas pUTI89-like plasmids appear to be mostly associated with humans via ST95 and ST131 clades A and B (Cusumano et al. 2010; Stephens et al. 2017; McKinnon et al. 2018; Moran and Hall 2018; Cummins et al. 2019; Li et al. 2021; Cummins et al. 2022). Our data supports the human association of pUTI89-like plasmids within ST127 and extends this range to companion animals. The absence of ColV plasmids is also consistent with very few livestock origin isolates present in the collection. Furthermore, whilst ColV plasmids encode numerous genes with purported roles in both fitness and pathogenicity, curing of pUTI89 was demonstrated to have no effect on fitness factors such as growth, type 1 pilus expression or biofilm formation but resulted in a significant reduction in pathogenicity within a mouse model of UTI (Cusumano et al. 2010). This supports the role of pUTI89-like plasmids in enhancing virulence in a proportion of ST127 isolates, particularly the dominant BAP4 cluster but does not support a plasmid-mediated fitness advantage in non-pathogenic sites. The association of pUTI89 with low carriage of ARGs in our collection has been previously noted and it was hypothesised that pUTI89-like plasmids may exert some form of exclusion on other mobile genetic elements (Stephens et al. 2017). However, ST127 isolates with pUTI89, class 1 integrons and multiple ARGs in our collection indicates any such mechanism is by no means absolute.

Although the epidemiological background of our collection was reasonably diverse, our study is limited by its comparatively small sample size (n=299) and the relatively limited availability of sequenced *E. coli* genomes prior to the 1990s. Additionally, *E. coli* genomes originating from some source niches and geographic locations were considerably under-represented in the publicly available databases, from which a large portion of the sequences used in this study were sourced. Nonetheless, the dominance of human and companion animal sequences reflects reports of ST127 in the available literature. Furthermore, the clear population structure, global distribution of isolates, diversity of AMR genotypes and characteristic features of the large BAP4 cluster tend to suggest we have observed most of the genetic diversity that could conceivably have been examined at this time.

In conclusion, this study supports previous work that indicates *E. coli* ST127 is a globally disseminated extraintestinal pathogen of humans and companion animals characterised by high VAG carriage and a lower prevalence of ARGs. In addition to previous literature, we have identified a major cluster of ST127 with a) numerous close linkages between human and companion animal isolates, b) carriage of pUTI89-like virulence plasmids and c) a repertoire of gene alleles that might provide clues to selective factors involved in its emergence. Future studies on ST127 could seek to elucidate the role and significance of pUTI89-like plasmids and genes associated with the dominant BAP4 cluster.

Funding

This research is funded by the Australian Government Research Training Programme and the Australian Center for Genomic Epidemiological Microbiology (AusGEM), a collaborative partnership between the New Souths Department of Primary Industries and the University of Technology Sydney.

Supplementary Figures

Fig S1. Presence/absence of ARGs mapped to core gene phylogeny
Fig S2. Presence/absence of VAGs mapped to core gene phylogeny
Fig S3. Presence/absence of plasmid-associated genes mapped to core gene phylogeny

Supplementary Tables

Table S1. Metadata, accession numbers and gene screening results for 299 ST127 isolates used in this study
Table S2. Integron co-carriage data; summary of assembly contigs containing the *intI1* gene in conjunction with ARGs

CRediT authorship contribution statement

Paarthiphan Elankumaran: Formal analysis, Investigation, Data curation, Writing – original draft. **Glenn F. Browning:** Investigation, Data curation, Project administration. **Marc S. Marenda:** Investigation, Data curation, Project administration. **Cameron J. Reid:** Conceptualization, Methodology, Software, Validation, Formal analysis, Writing – review & editing, Visualization, Supervision. **Steven P. Djordjevic:** Conceptualization, Resources, Writing – review & editing, Supervision, Project administration, Funding acquisition.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Steven Djordjevic reports financial support was provided by AusGEM.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.crmicr.2022.100106.

References

Alghoribi, M.F., Gibreel, T.M., Dodgson, A.R., Beatson, S.A., Upton, M., 2014. Galleria mellonella infection model demonstrates high lethality of ST69 and ST127 uropathogenic E. coli. PLoS One 9 (7), e101547.
Alghoribi, M.F., Gibreel, T.M., Farnham, G., Al Johani, S.M., Balkhy, H.H., Upton, M., 2015. Antibiotic-resistant ST38, ST131 and ST405 strains are the leading uropathogenic Escherichia coli clones in Riyadh, Saudi Arabia. J. Antimicrob. Chemother. 70 (10), 2757–2762.

Alkeskas, A., Ogrodzki, P., Saad, M., Masood, N., Rhoma, N.R., Moore, K., Farbos, A., Paszkiewicz, K., Forsythe, S., 2015. The molecular characterisation of Escherichia coli K1 isolated from neonatal nasogastric feeding tubes. BMC Infectious Diseases 15 (1), 449.

Banerjee, R., Johnston, B., Lohse, C., Chattopadhyay, S., Tchesnokova, V., Sokurenko, E. V., Johnson, J.R., 2013. The clonal distribution and diversity of extraintestinal Escherichia coli isolates vary according to patient characteristics. Antimicrob. Agents Chemother. 57 (12), 5912–5917.

Beyrouthy, R., Robin, F., Cougnoux, A., Dalmasso, G., Darfeuille-Michaud, A., Mallat, H., Dabboussi, F., Hamze, M., Bonnet, R., 2013. Chromosome-mediated OXA-48 carbapenemase in highly virulent Escherichia coli. J. Antimicrob. Chemother. 68.

Bogema, D.R., McKinnon, J., Liu, M., Hitchick, N., Miller, N., Venturini, C., Iredell, J., Darling, A.E., Chowdury, P.Roy, Djordjevic, S.P., 2020. 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 6 (1).

Bonnet, R., Beyrouthy, R., Haenni, M., Nicolas-Chanoine, M.H., Dalmasso, G., Madec, J. Y., 2021. Host Colonization as a Major Evolutionary Force Favoring the Diversity and the Emergence of the Worldwide Multidrug-Resistant Escherichia coli ST131. mBio 12 (4), e0145121.

Bourne, J.A., Chong, W.L., Gordon, D.M., 2019. 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 14 (3), e0212867.

Brynildsrud, O., Bohlin, J., Scheffer, L., Eldholm, V., 2016. Rapid scoring of genes in microbial pan-genome-wide association studies with Scoary. Genome Biol. 17 (1), 238.

Carattoli, A., Hasman, H., 2020. PlasmidFinder and In Silico pMLST: Identification and Typing of Plasmid Replicons in Whole-Genome Sequencing (WGS). Methods Mol. Biol. 2075, 285–294.

Carattoli, A., Zankari, E., García-Fernández, A., Voldby Larsen, M., Lund, O., Villa, L., Møller Aarestrup, F., Hasman, H., 2014. In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. Antimicrob. Agents Chemother. 58 (7), 3895–3903.

Chen, L., Zheng, D., Liu, B., Yang, J., Jin, Q., 2016. VFDB 2016: hierarchical and refined dataset for big data analysis–10 years on. Nucleic. Acids. Res. 44 (D1), D694–D697.

Ciesielczuk, H., Jenkins, C., Chattaway, M., Doumith, M., Hope, R., Woodford, N., Wareham, D.W., 2016. Trends in ExPEC serogroups in the UK and their significance. European journal of clinical microbiology & infectious diseases: official publication of the European Society of Clinical Microbiology 35 (10), 1661–1666.

Croxall, G., Hale, J., Weston, V., Manning, G., Cheetham, P., Achtman, M., McNally, A., 2011. Molecular epidemiology of extraintestinal pathogenic Escherichia coli isolates from a regional cohort of elderly patients highlights the prevalence of ST131 strains with increased antimicrobial resistance in both community and hospital care settings. J. Antimicrob. Chemother. 66 (11), 2501–2508.

Cummins, M.L., Reid, C.J., Djordjevic, S.P., 2022. F Plasmid Lineages in Escherichia coli ST95: Implications for Host Range, Antibiotic Resistance, and Zoonoses. mSystems 7 (1).

Cummins, M.L., Reid, C.J., Chowdhury, P.Roy, Bushell, R.N., Esbert, N., Tivendale, K.A., Noormohammadi, A.H., Islam, S., Marenda, M.S., Browning, G.F., Markham, P.F., Djordjevic, S.P., 2019. Whole genome sequence analysis of Australian avian pathogenic Escherichia coli that carry the class 1 integrase gene. Microb Genom 5 (2).

Cusumano, C.K., Hung, C.S., Chen, S.L., Hultgren, S.J., 2010. Virulence plasmid harbored by uropathogenic Escherichia coli functions in acute stages of pathogenesis. Infect. Immun. 78 (4), 1457–1467.

Darling, A.E., McKinnon, J., Worden, P., Santos, J., Charles, I.G., Chowdhury, P.R., Djordjevic, S.P., 2014. A draft genome of Escherichia coli sequence type 127 strain 2009-46. Gut Pathogens 6 (1), 32.

DebRoy, C., Sidhu, M.S., Sarker, U., Jayarao, B.M., Stell, A.L., Bell, N.P., Johnson, T.J., 2010. Complete sequence of pEC14_114, a highly conserved IncFIB/FIIA plasmid associated with uropathogenic Escherichia coli cystitis strains. Plasmid 63 (1), 53–60.

Fibke, C.D., Croxen, M.A., Geum, H.M., Glass, M., Wong, E., Avery, B.P., Daignault, D., Mulvey, M.R., Reid-Smith, R.J., Parmley, E.J., Portt, A., Boerlin, P., Manges, A.R., 2019. Genomic Epidemiology of Major Extraintestinal Pathogenic Escherichia coli Lineages Causing Urinary Tract Infections in Young Women Across Canada. Open forum infectious diseases 6 (11), ofz431.

Flament-Simon, S.C., Toro, M., Garcia, V., Blanco, J.E., Blanco, M., Alonso, M.P., Goicoa, A., Diaz-Gonzalez, J., Nicolas-Chanoine, M.H., Blanco, J., 2020. Molecular Characteristics of Extraintestinal Pathogenic E. coli (ExPEC), Uropathogenic E. coli (UPEC), and Multidrug Resistant E. coli Isolated from Healthy Dogs in Spain. Whole Genome Sequencing of Canine ST372 Isolates and Comparison with Human Isolates Causing Extraintestinal Infections. Microorganisms 8 (11).

Gaio, D., J. To, M. Liu, L. Monahan, K. Anantanawat and A.E. Darling (2019). "Hackflex: low cost Illumina sequencing library construction for high sample counts." bioRxiv: 779215.

Gibreel, T.M., Dodgson, A.R., Cheesbrough, J., Fox, A.J., Bolton, F.J., Upton, M., 2012. Population structure, virulence potential and antibiotic susceptibility of uropathogenic Escherichia coli from Northwest England. J. Antimicrob. Chemother. 67 (2), 346–356.

Gomi, R., Matsuda, T., Matsumura, Y., Yamamoto, M., Tanaka, M., Ichiyama, S., Yoneda, M., 2017. Whole-Genome Analysis of Antimicrobial-Resistant and Extraintestinal Pathogenic Escherichia coli in River Water. Appl. Environ. Microbiol. 83 (5).

Hastak, P., Cummins, M.L., Gottlieb, T., Cheong, E., Merlino, J., Myers, G.S.A., Djordjevic, S.P., Chowdhury, P.Roy, 2020. Genomic profiling of Escherichia coli isolates from bacteraemia patients: a 3-year cohort study of isolates collected at a Sydney teaching hospital. Microb Genom 6 (5).

Hertz, F.B., Nielsen, J.B., Schønning, K., Littauer, P., Knudsen, J.D., Løbner-Olesen, A., Frimodt-Møller, N., 2016. Population structure of Drug-Susceptible, -Resistant and ESBL-producing Escherichia coli from Community-Acquired Urinary Tract Infections. BMC Microbiol. 16 (1), 63.

Horner, C., Fawley, W., Morris, K., Parnell, P., Denton, M., Wilcox, M., 2014. Escherichia coli bacteraemia: 2 years of prospective regional surveillance (2010-12). J. Antimicrob. Chemother. 69 (1), 91–100.

Ingle, D.J., Valcanis, M., Kuzevski, A., Tauschek, M., Inouye, M., Stinear, T., Levine, M. M., Robins-Browne, R.M., Holt, K.E., 2016. 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. Microbial genomics 2 (7), e000064.

Jia, B., Raphenya, A.R., Alcock, B., Waglechner, N., Guo, P., Tsang, K.K., Lago, B.A., Dave, B.M., Pereira, S., Sharma, A.N., Doshi, S., Courtot, M., Lo, R., Williams, L.E., Frye, J.G., Elsayegh, T., Sardar, D., Westman, E.L., Pawlowski, A.C., Johnson, T.A., Brinkman, F.S., Wright, G.D., McArthur, A.G., 2017. CARD 2017: expansion and model-centric curation of the comprehensive antibiotic resistance database. Nucleic. Acids. Res. 45 (D1), D566–D573.

Johnson, J., Johnston, B., Clabots, C., Kuskowski, M., Roberts, E., DebRoy, C., 2008a. Virulence Genotypes and Phylogenetic Background of Escherichia coli Serogroup O6 Isolates from Humans, Dogs, and Cats. J. Clin. Microbiol. 46, 417–422.

Johnson, J.R., Clabots, C., Kuskowski, M.A., 2008b. Multiple-host sharing, long-term persistence, and virulence of Escherichia coli clones from human and animal household members. J. Clin. Microbiol. 46 (12), 4078–4082.

Johnson, J.R., O'Bryan, T.T., Low, D.A., Ling, G., Delavari, P., Fasching, C., Russo, T.A., Carlino, U., Stell, A.L., 2000. Evidence of commonality between canine and human extraintestinal pathogenic Escherichia coli strains that express papG allele III. Infect. Immun. 68 (6), 3327–3336.

Johnson, J.R., Owens, K., Gajewski, A., Clabots, C., 2008. Escherichia coli colonization patterns among human household members and pets, with attention to acute urinary tract infection. J. Infect. Dis. 197 (2), 218–224.

Johnson, T.J., 2021. Role of Plasmids in the Ecology and Evolution of "High-Risk" Extraintestinal Pathogenic Escherichia coli Clones. EcoSal Plus 9 (2).

Jolley, K.A., Maiden, M.C.J., 2010. BIGSdb: Scalable analysis of bacterial genome variation at the population level. BMC Bioinf. 11 (1), 595.

Kallonen, T., Brodrick, H.J., Harris, S.R., Corander, J., Brown, N.M., Martin, V., Peacock, S.J., Parkhill, J., 2017. Systematic longitudinal survey of invasive Escherichia coli in England demonstrates a stable population structure only transiently disturbed by the emergence of ST131. Genome Res. 27 (8), 1437–1449.

Kidsley, A., O'Dea, M., Saputra, S., Jordan, D., Johnson, J., Gordon, D., Turni, C., Djordjevic, S., Abraham, S., Trott, D., 2020a. Genomic Analysis of Phylogenetic Group B2 Extraintestinal Pathogenic E. coli Causing Infections in Dogs in Australia. Vet. Microbiol. 248, 108783.

Kidsley, A.K., O'Dea, M., Ebrahimie, E., Mohammadi-Dehcheshmeh, M., Saputra, S., Jordan, D., Johnson, J.R., Gordon, D., Turni, C., Djordjevic, S.P., Abraham, S., Trott, D.J., 2020b. Genomic analysis of fluoroquinolone-susceptible phylogenetic group B2 extraintestinal pathogenic Escherichia coli causing infections in cats. Vet. Microbiol. 245, 108685.

Kornberg, H.L., 1966. The role and control of the glyoxylate cycle in Escherichia coli. Biochem. J. 99 (1), 1–11.

Koster, J., Rahmann, S., 2012. Snakemake–a scalable bioinformatics workflow engine. Bioinformatics 28 (19), 2520–2522.

LeCuyer, T.E., Byrne, B.A., Daniels, J.B., Diaz-Campos, D.V., Hammac, G.K., Miller, C.B., Besser, T.E., Davis, M.A., 2018. Population Structure and Antimicrobial Resistance of Canine Uropathogenic <span class="named-content genus-species" id="named-content-1">Escherichia coli</span&gt. J. Clin. Microbiol. 56 (9) e00788-00718.

Li, D., Wyrsch, E.R., Elankumaran, P., Dolejska, M., Marenda, M.S., Browning, G.F., Bushell, R.N., McKinnon, J., Chowdhury, P.R., Hitchick, N., Miller, N., Donner, E., Drigo, B., Baker, D., Charles, I.G., Kudinha, T., Jarocki, V.M., Djordjevic, S.P., 2021. Genomic comparisons of Escherichia coli ST131 from Australia. Microb Genom 7 (12).

Liu, C., Stegger, M., Aziz, M., Johnson, T., Schmidlin, K., Nordstrom, L., Gauld, L., Weaver, B., Rolland, D., Statham, S., Horwinski, J., Sariya, S., Davis, G., Sokurenko, E., Keim, P., Johnson, J., Price, L., 2018. Escherichia coli ST131- H 22 as a Foodborne Uropathogen. mBio 9.

Lüthje, P., Brauner, A., 2014. Chapter Seven - Virulence Factors of Uropathogenic E. coli and Their Interaction with the Host. Advances in Microbial Physiology. R. K. Poole 65, 337–372. Academic Press.

Maluta, R.P., Stella, A.E., Riccardi, K., Rigobelo, E.C., Marin, J.M., Carvalho, M.B., de Ávila, F.A., 2012. Phenotypical characterization and adhesin identification in Escherichia coli strains isolated from dogs with urinary tract infections. Braz J Microbiol 43 (1), 375–381.

Manges, A.R., Geum, H.M., Guo, A., Edens, T.J., Fibke, C.D., Pitout, J.D.D., 2019. Global Extraintestinal Pathogenic Escherichia coli (ExPEC) Lineages. Clin. Microbiol. Rev. 32 (3).

McKinnon, J., Chowdhury, P.Roy, Djordjevic, S.P., 2018. Genomic analysis of multidrug-resistant Escherichia coli ST58 causing urosepsis. Int. J. Antimicrob. Agents 52 (3), 430–435.

Melendez, D., Roberts, M.C., Greninger, A.L., Weissman, S., No, D., Rabinowitz, P., Wasser, S., 2019. 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. 74 (8), 2176–2180.

Moran, R.A., Hall, R.M., 2018. Evolution of Regions Containing Antibiotic Resistance Genes in FII-2-FIB-1 ColV-Colla Virulence Plasmids. Microb. Drug Resist. 24 (4), 411–421.
Nesporova, K., Wyrsch, E.R., Valcek, A., Bitar, I., Chaw, K., Harris, P., Hrabak, J., Literak, I., Djordjevic, S.P., Dolejska, M., 2020. Escherichia coli Sequence Type 457 Is an Emerging Extended-Spectrum-β-Lactam-Resistant Lineage with Reservoirs in Wildlife and Food-Producing Animals. Antimicrob. Agents Chemother. 65 (1).
Nguyen, L.T., Schmidt, H.A., von Haeseler, A., Minh, B.Q., 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. 32 (1), 268–274.
Nowak, K., Fahr, J., Weber, N., Lübke-Becker, A., Semmler, T., Weiss, S., Mombouli, J.-V., Wieler, L.H., Guenther, S., Leendertz, F.H., Ewers, C., 2017. Highly diverse and antimicrobial susceptible Escherichia coli display a naïve bacterial population in fruit bats from the Republic of Congo. PLoS One 12 (7), e0178146.
Page, A.J., Cummins, C.A., Hunt, M., Wong, V.K., Reuter, S., Holden, M.T., Fookes, M., Falush, D., Keane, J.A., Parkhill, J., 2015. Roary: rapid large-scale prokaryote pan genome analysis. Bioinformatics 31 (22), 3691–3693.
Page, A.J., Taylor, B., Delaney, A.J., Soares, J., Seemann, T., Keane, J.A., Harris, S.R., 2016. SNP-sites: rapid efficient extraction of SNPs from multi-FASTA alignments. Microb Genom 2 (4), e000056.
Petty, N.K., Ben Zakour, N.L., Stanton-Cook, M., Skippington, E., Totsika, M., Forde, B. M., Phan, M.D., Gomes Moriel, D., Peters, K.M., Davies, M., Rogers, B.A., Dougan, G., Rodriguez-Baño, J., Pascual, A., Pitout, J.D., Upton, M., Paterson, D.L., Walsh, T.R., Schembri, M.A., Beatson, S.A., 2014. Global dissemination of a multidrug resistant Escherichia coli clone. Proc. Natl. Acad. Sci. U. S. A. 111 (15), 5694–5699.
Reid, C.J., Blau, K., Jechalke, S., Smalla, K., Djordjevic, S.P., 2020. Whole Genome Sequencing of Escherichia coli From Store-Bought Produce. Frontiers in microbiology 10, 3050.
Reid, C.J., McKinnon, J., Djordjevic, S.P., 2019. 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 5 (9).
Reid, C.J., Chowdhury, P.Roy, Djordjevic, S.P., 2015. Tn6026 and Tn6029 are found in complex resistance regions mobilised by diverse plasmids and chromosomal islands in multiple antibiotic resistant Enterobacteriaceae. Plasmid 80, 127–137.
Riley, L.W., 2014. Pandemic lineages of extraintestinal pathogenic Escherichia coli. Clin. Microbiol. Infect. 20 (5), 380–390.
Salipante, S.J., Roach, D.J., Kitzman, J.O., Snyder, M.W., Stackhouse, B., Butler-Wu, S. M., Lee, C., Cookson, B.T., Shendure, J., 2015. Large-scale genomic sequencing of extraintestinal pathogenic Escherichia coli strains. Genome Res. 25 (1), 119–128.
Sarkar, S., Ulett, G.C., Totsika, M., Phan, M.-D., Schembri, M.A., 2014. Role of Capsule and O Antigen in the Virulence of Uropathogenic Escherichia coli. PLoS One 9 (4), e94786.
Seemann, T., 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30 (14), 2068–2069.
Siguier, P., Perochon, J., Lestrade, L., Mahillon, J., Chandler, M., 2006. ISfinder: the reference centre for bacterial insertion sequences. Nucleic. Acids. Res. 34 (Database issue), D32–D36.
Smalla, K., Cook, K., Djordjevic, S.P., Klümper, U., Gillings, M., 2018. Environmental dimensions of antibiotic resistance: assessment of basic science gaps. FEMS Microbiol. Ecol. 94 (12).
Soysal, N., Mariani-Kurkdjian, P., Smail, Y., Liguori, S., Gouali, M., Loukiadis, E., Fach, P., Bruyand, M., Blanco, J., Bidet, P., Bonacorsi, S., 2016. Enterohemorrhagic Escherichia coli Hybrid Pathotype O80:H2 as a New Therapeutic Challenge. Emerg. Infect. Dis. 22 (9), 1604–1612.
Stephens, C.M., Adams-Sapper, S., Sekhon, M., Johnson, J.R., Riley, L.W., 2017. Genomic Analysis of Factors Associated with Low Prevalence of Antibiotic Resistance in Extraintestinal Pathogenic Escherichia coli Sequence Type 95 Strains. mSphere 2 (2).
Tonkin-Hill, G., Lees, J.A., Bentley, S.D., Frost, S.D.W., Corander, J., 2019. Fast hierarchical Bayesian analysis of population structure. Nucleic. Acids. Res. 47 (11), 5539–5549.
Ulleryd, P., Sandberg, T., Scheutz, F., Clabots, C., Johnston, B.D., Thuras, P., Johnson, J. R., 2015. Colonization with Escherichia coli Strains among Female Sex Partners of Men with Febrile Urinary Tract Infection. J. Clin. Microbiol. 53 (6), 1947–1950.
Ward, Doyle V., Scholz, M., Zolfo, M., Taft, Diana H., Schibler, Kurt R., Tett, A., Segata, N., Morrow, Ardythe L., 2016. Metagenomic Sequencing with Strain-Level Resolution Implicates Uropathogenic E. coli in Necrotizing Enterocolitis and Mortality in Preterm Infants. Cell Rep. 14 (12), 2912–2924.
Wyrsch, E.R., Chowdhury, P.R., Jarocki, V.M., Brandis, K.J., Djordjevic, S.P., 2020. Duplication and diversification of a unique chromosomal virulence island hosting the subtilase cytotoxin in Escherichia coli ST58. Microb Genom 6 (6).
Yamaji, R., Rubin, J., Thys, E., Friedman, C.R., Riley, L.W., 2018. Persistent Pandemic Lineages of Uropathogenic Escherichia coli in a College Community from 1999 to 2017. J. Clin. Microbiol. 56 (4).
Zankari, E., Allesøe, R., Joensen, K.G., Cavaco, L.M., Lund, O., Aarestrup, F.M., 2017. PointFinder: a novel web tool for WGS-based detection of antimicrobial resistance associated with chromosomal point mutations in bacterial pathogens. J. Antimicrob. Chemother. 72 (10), 2764–2768.
Zhou, Z., Alikhan, N.F., Mohamed, K., Fan, Y., Agama Study, G., Achtman, M., 2020. The EnteroBase user's guide, with case studies on Salmonella transmissions, Yersinia pestis phylogeny, and Escherichia core genomic diversity. Genome Res. 30 (1), 138–152.
Zurfluh, K., Tasara, T., Stephan, R., 2016. Full-Genome Sequence of Escherichia coli K-15KW01, a Uropathogenic E. coli B2 Sequence Type 127 Isolate Harboring a Chromosomally Carried blaCTX-M-15 Gene. Genome Announc. 4 (5) e00927-00916.