An emerging lineage of uropathogenic extended spectrum β-lactamase Escherichia coli ST127

An emerging lineage of uropathogenic extended spectrum beta-lactamase *Escherichia coli* ST127

Rory Cave[1], Mary M Ter-Stepanyan[2,3], Nune Kotsinyan[4], Hermine V Mkrtchyan[1]

[1]School of Biomedical Sciences, University of West London, London, W5 5RF, UK

[2]Yerevan State Medical University after M. Heratsi, Faculty of Public Health, Department of Epidemiology, Yerevan, 2 Koryun street, 0025, Republic of Armenia

[3]Research Center of Maternal and child health protection, 22 Mashtots Ave., Yerevan, 0002, Armenia

[4]National Centre for Disease Control and Prevention, 12 Mkhitar Heratsi street, Yerevan, Armenia, 0025

Corresponding author: Hermine Mkrtchyan (Hermine.mkrtchyan@uwl.ac.uk)

Abstract

Uropathogenic *Escherichia coli* (UPEC) is one of the most common causes of urinary tract infections. Here we report for the first time the whole genome sequencing (WGS) and analysis of four extended-spectrum beta-lactamase (ESBL) UPEC sequence type (ST) 127 isolates recovered from patients in five hospitals in Armenia between January to August 2019. Phylogenetic comparison revealed that our isolates were closely related to each other by the core and accessory genomes despite being isolated from different regions and hospitals in Armenia. We identified unique genes in our isolates and in a closely related isolate recovered in France. The unique genes (hemolysin E virulence gene, lactate utilisation operon *lutABC* and endonuclease restriction modification operon *hsdMSR*) were identified in three separate genomic regions adjacent to prophage genes, including one region containing the TonB-dependent iron siderophore receptor gene *ireA*, which was only found in 5 other ST127 isolates from the European Nucleotide Archive (ENA). We further identified that these isolates possessed unique virulence and metabolic genes and harboured antibiotic resistance genes, including ESBL genes $bla_{\text{CTX-M-3}}$ (n=3), $bla_{\text{CTX-M-236}}$ (n=1) and $bla_{\text{TEM-1}}$ (n=1), in addition to a quinolone resistance protein gene *qnrD1* (n=1), which was absent in ST127 isolates obtained from ENA. Moreover, a plasmid replicon gene IncI2 (n=1) was unique to ARM88 of the Armenian isolates. Our findings demonstrate that at the time of this study *E. coli* ST127 was a cause of urinary tract infections in patients in different regions of Armenia, with a possibility of cross-country transmission between Armenia and France.

Importance

Whole genome sequencing studies of pathogens causing infectious diseases are seriously lacking in Armenia, hampering global efforts to track, trace and contain infectious disease outbreaks. In this study, we report for the first-time whole genome sequencing and analysis of ESBL UPEC ST127 isolates recovered from hospitalised patients in Armenia and compare them with other *E. coli* ST127 retrieved from the ENA.

We found close genetic similarities of the Armenian isolates indicating that *E. coli* ST127 was potentially a dominant lineage causing urinary tract infections in Armenia. Furthermore, we identified unique genes horizontally acquired in the Armenian and French isolates clusters that were absent in other ST127 isolates obtained from the ENA.

Our findings highlight a possible cross-country transmission between Armenia and France and that implementation of the WGS surveillance could contribute to the global efforts in tackling antibiotic resistance as bacteria carrying AMR genes do not recognise borders.

Introduction

Escherichia coli is one of the most common causes of urinary tract infections, estimated to affect about 150 million people globally (1). New emerging lineages present a major challenge for healthcare settings due to increased resistance to multiple antibiotics (Kot, 2019), contributing to high number of hospitalisations and associated costs. Uropathogenic *E. coli* (UPEC) pathotype is the primary cause of UTI infections and is a part of the broader pathotype designated as Extraintestinal Pathogenic *E. coli* (ExPEC), known to cause infections in the bloodstream and other non-intestinal sites (3). Further members of the ExPEC pathotype include neonatal meningitis *E. coli* (NMEC), sepsis-associated *E. coli* (SEPEC), and avian pathogenic *E. coli* (APEC) (4). One of the new emerging ExPEC pandemic genotypes of interest is sequence type (ST) 127, known for its high virulence but low antibiotic resistance potential compared to other pandemic ExPEC sequence types (5–7). The acquisition of many virulence and antibiotic resistance determents by UPEC isolates is through horizontal transfer of plasmids and prophages that can integrate into the bacterial chromosome (8–10). These acquired genes are only maintained long-term if they demonstrate positive evolutional adaption with little fitness cost under nonselective conditions (11). An example of this is the plasmid-borne CTX-M-type genes that encode for an extended-spectrum β-lactamases (ESBL), which gives the bacteria resistance towards expanded spectrum cephalosporins and monobactams (12). Over the past decade the prevalence of ESBL producing *E. coli* has increased globally (13, 13–15). However, ESBL producing UPEC ST127 isolates are rarely reported in the literature (5, 16, 17)

Our understanding of UPEC infections has greatly improved due to the power of whole genome sequencing (WGS) analysis to trace the source of infection, their transmission routes, as well as identifying genes involved in virulence and antibiotic resistance for implementing improved infection control practices (6, 18–20). However, studies reporting WGS are limited to some regions of the world (i.e. UK, EU, North America), whereas studies from low- middle-income countries are fragmented (21, 22). There have been very limited studies utilising WGS in Armenia, an upper-middle-income country (23) (according to the World Bank in 2022), with none reported for *E. coli*. Previously, we reported the occurrence of diverse MRSA genotypes (24) and provided insights into the genomic background and phylogenetic origins of MRSA isolates in Armenia (25). In this study, we report for the first time the genetic features of ESBL UPEC ST127 isolates recovered from hospitalised patients in Armenia using WGS analysis and compare them with 168 other *E. coli* ST127 recovered from multiple sources (available at the European Nucleotide Archive (ENA)) and identify the mode of horizontal transfer of the unique genes found only in the Armenian isolates.

Methods

Bacterial isolation and identification

Twelve *E. coli* isolates have been received from Medical Microbiology laboratories of five hospitals in Armenia between January and August 2019. All isolates were recovered from urine specimen of hospitalised patients. The isolates were speciated as *E. coli* using a matrix-assisted laser desorption ionization time flight mass-spectroscopy (MALDI-TOF-MS) as described previously (26). Four out of 12 isolates that have been identified as *E. coli* ST127 were selected for the purposes of this study.

Antibiotic Susceptibility Testing

All twelve isolates were tested for their antibiotic susceptibility to a panel of eleven antibiotics, including Ampicillin (10mg), Piperacillin/Tazobactam (30/6mg), Amoxicillin and Clavulanic Acid (20/10mg), Ceftazidime (10mg), Cefepime (30mg), Norfloxacin (10mg), Levofloxacin (5mg), Amikacin (30mg), Imipenem (10mg), Meropenem (10mg), Chloramphenicol (30mg) (Mast Group,

Merseyside, UK) using a disk diffusion method according to the European Committee on Antimicrobial Susceptibility Testing protocol (27). Antibiotics chosen were the most frequently used in clinical settings in Armenia. *E. coli* isolates were identified as 'ESBL producing' upon confirming their resistance to cefepime and ceftazidime antibiotics.

Genome sequencing and Assembly

All *E. coli* isolates were whole genome sequenced using the Illumina HiSeq platform. However, for the purposes of this study WGS analysis were conducted for only four isolates belonging to ST127. Genomic DNA was extracted using the TIANamp Bacteria DNA kit (Tiangen, China) and paired-end sequencing libraries were constructed using Nextera XT DNA Sample Preparation kits or TruSeq DNA HT Sample Prep Kit (Illumina, USA) following the manufacturer's instruction.

The quality of short-reads was analysed using fastQC, and low quality reads were trimmed using the software Trimmomatic (28). The trimmed reads we De-novo assembled using SPAdes (29).

Genome selection for phylogenetic and genomic comparison

To conduct a comparative genomic analysis of *E. coli* ST127 isolates recovered in this study, 168 draft *E. coli* ST127 genomes (Table S1) were obtained from the ENA database (accessed January 2022). The selection criteria of the isolates included: date and source of isolation and country of origin. To select the ST127 genomes only among all 12 *E.coli* isolates recovered in this study and those retrieved from ENA database, we screened the draft *E. coli* genomes using mlst (https://github.com/tseemann/mlst accessed July 2021) using the Achtman typing scheme from pubMLST (accessed January 2022).

Phylogenetic analysis

To construct a core SNP maximum-likelihood (ML) phylogenetic tree of *E. coli* ST127 isolates, we first aligned all isolates against the reference *E. coli* ST127 genome EC536 (accession no. NC_008253.1) using the Parsnps alignment software from the Harvest suite (30). Then, recombination was removed from aligned sequences using Gubbins, and a phylogenetic tree was constructed from the recombination-free alignment using IQtree v2.1.2 (31). To select the best model for phylogenetic tree construction, we used ModelFinder and set ultrafast bootstrap replication to 1,000 (32, 33). Finally, the phylogenetic tree was visualised using iTOL (34). The SNP -distance between isolates core genome was worked out from the recombination-free alignment using SNP-dist software (https://github.com/tseemann/snp-dists).

A Bayesian dated maximum clade credibility (MCC) tree was reconstructed using BEAST v1.10.4 (35). A GTR empirical substitution model as determined by ModelFinder was used for the Bayesian analysis set with a Bayesian skyline strict clock model with the MCC chain set to 100 million. Two independent runs trees were combined using LogCombiner with a 10% burnin and an MCC tree constructed from combined trees using TreeAnnotator. MCC tree was visualised using the software FigTree (http://tree.bio.ed.ac.uk/software/figtree/).

Genome annotation

E. coli ST127 isolates genomes were annotated using Prokka (36). Virulent gene marker and plasmid replicons were screened using Abricate software (https://github.com/tseemann/abricate) in conjunction

with the Virulencefinder and VFDB database combined for virulence detection and Palsmidfinder database to identify plasmid replicons (accessed January 2022) (37–39). Antibiotic resistance gene/mutation was screened using the Resistance Gene Identifier software in conjugation with the Comprehensive Resistant Antibiotic Resistance Database (accessed January 2022) (40). Isolates were serotyped by ECtyper (41).

Pangenome and unique gene analysis

The pangenome combining both the Armenian and ENA *E. coli* ST127 isolates was constructed using Roary with the blastp percentage identity cutoff set at 90% (42). Unique genes within different *E. coli* ST127 groups were identified using Scoary (43). Further annotation of unique genes and genes within their loci was further annotated by blasting against the uniport and NCBI nucleotide collection. MAFFT was used to identify individual protein percentage similarities (44). Genomic region comparison was visualised using mauve v 2.4.0 (45).

MOB-suit was used to reconstruct plasmid sequence from draft genome to determine if unique genes were plasmid-borne (46). Integrative and conjugative element genomic regions were detected using ICEfinder using the ICEberg 2.0 database (47) Prophage hunter was to identify the prophage region in the draft genome to determine if unique genes were within the prophage insertion region (48). The horizontal gene transfer regions containing these unique genes were compared to other *E. coli* isolates by sequence similarity search using NCBI blast against the ENA sequence archive (49).

R programs

Grouping isolates into subclusters within the SNP core whole genome alignment was conducted by the hierarchical clustering R package "RhierBAPS" (50). In addition, a hierarchy cluster heatmap was constructed of antibiotic resistant gene/mutation, accessory genes and plasmid replicon using the R package pheatmap (https://cran.r-project.org/web/packages/pheatmap/index.html)

Data Availability

The short-read data were deposited in the ENA, under the study PRJEB51925. Individual sequence file accession numbers are included in the Supplementary Table 2 (Table S2).

Results

Isolates and Antibiotic Susceptibility testing

4 out the twelve sequenced *E. coli* isolates (Table S3) belonged to ST127 (Table 1) and were resistant to 7 (n=1), 6 (n=1) and 5 (n=2) of antibiotics tested. All four isolates were resistant to cephalosporin antibiotics cefepime, ceftazidime and aminopenicillins antibiotic ampicillin. In addition, 3 isolates (ARM64, ARM 75 and ARM88) were resistant to the β-lactam antibiotic amoxicillin-clavulanic acid, 2 isolates (ARM64 and ARM66) had intermediate resistance to the β-lactam antibiotic piperacillin-tazobactam, one isolate (ARM88) was resistant to the aminoglycoside antibiotic amikacin fluoroquinolone antibiotics norfloxacin, and levofloxacin (ARM75) and one isolate had intermediate resistant to the carbapenems antibiotic imipenem (ARM66). All four isolates were sensitive to β-lactam antibiotics meropenem and were also sensitive to chloramphenicol.

Phylogenetic analysis of *E. coli* ST127

The short reads of four *E. coli* ST127 in this study were aligned against 168 *E. coli* ST127 genomes obtained from the ENA archive that were previously recovered from 15 countries and 11 different animal sources and belonged to three different serotypes (O6:H31 (n=161), O75:H31 (n=3), -:H321(n=8)) (Table 2). A maximum-likelihood phylogenetic tree of the core genome revealed that ST127 could be split into 3 main clades with no particular clade having isolates that are only found from one particular country or source (Figure 1A). All O75:H31 serotype isolates cluster together in clade B. Further detailed analyses of the clusters (using heirBAPS) revealed that ST127 isolates could be further grouped into 5 distinct clusters. *E. coli* ST127 isolates recovered in our study belonged to clade C BAP2 and were phylogenetically closely related to each other, with a maximum of 33 SNP differences between the isolates (Table S4). Moreover, all Armenian ST127 isolates were phylogenetically closely related to a French isolate (DABGLY01) that was recovered from a patient (urine sample) in 2015.

Maximum clade credibility (MCC) time-calibrated phylogeny tree of 171 *E. coli* ST127 isolates (excluding EC536 isolate due to lack of its date of isolation) was constructed using BEAST to determine the inferred date of divergence of the Armenian isolates and the French isolate (DABGLY01) obtained from the ENA archive to other ST127 isolates (Figure 2). The ST127 had a rate estimate of 9.18×10^{-5} substitution per site per year and inferred tree root date of 1503 (date confidence interval (CI) 1415 to 1605). The inferred divergence date of these isolates to their closest phylogenetically related isolate (The UK isolate DABBWT01 recovered from blood) was 1966 (date CI: 1956 to 1977). The most recent divergence date which both the Armenian and French isolate shared was 2000 (date CI:1995 to 2006) and the most recent divergence date between ARM75 and the closely phylogenetically related French isolate DABGLY01 was 2005 (date CI: 2000 to 2009).

Comparison of *E. coli* ST127 virulence genes

Overall, we identified 153 virulence genes in the *E. coli* ST127 isolates we analysed, of which 41 were ubiquities (Figure 1B). The Armenian isolates harboured 96 (ARM64), 95 (ARM75), 94 (ARM66) and 92 (ARM88) virulence genes, which are less than the mean average number (mean average=100, range 63 to 109) of virulence genes compared to other ST127 isolates analysed.

All four isolates in this study and a phylogenetically closely related French isolate DABGLY01 (<25 SNP difference between Armenian isolates within the core genomes), harboured the hemolysin E gene (*hylE*), which was not present in any of the other ST127 isolates. In addition, they also carried TonB-dependent iron siderophore receptor gene *ireA*, which was also found in five other ST127 isolates recovered from a dog (AATKBA01); a cat (AATMZD02), and human urine and faecal swabs (DABDVM01, DABMUI01, DABXBZ01); however, these isolates did not show any phylogenetic or geographical relationship to each other.

Unique genes and origin of horizontal gene transfer (HGT)

To further determine whether any of the genes were unique to the Armenian isolates as well as to the French isolate obtained from the ENA archive, we conducted a pangenome analysis of all *E. coli* ST127 isolates. The pangenome of the *E. coli* ST127 isolates consisted of 14,720 genes, of which 3,529 belonged to the core genome, whereas 11,191 to the accessory genome. The accessory genome

hierarchy clustering heatmap (Figure 3). Showed that the Armenian isolates shared many accessory genes with the DABGLY01.

Using Scoary, we identified 16 genes that were unique to all *E. coli* ST127 Armenian isolates but also to DABGLY01, recovered in France. 11 out of 16 genes identified had a known function (Table 3), including the hemolysin E toxin (*hylE*), lactate utilisation protein operon *lutABC*), Peptide ABC transporter substrate-binding protein, transposase *insD* for insertion element IS2, Inovirus Gp2 family protein, and the endonuclease restriction system operon (hsdMSR). We identified that 12 of these genes were located within the same locus next to a prophage integrase *intA* gene and included the *lutABC* operon gene and *oppA* gene. In addition, two genes located near the *ireA* gene were unique to the Armenian and DABGLY01 isolates were the *InsD* and the Inovirus Gp2 family protein gene. Moreover, we found one additional uncharacterised gene unique to Armenian isolates adjacent to the *hlyE* virulent gene.

To determine the origin of horizontal gene transfer of these unique genes found in Armenian isolates, we looked for insertion via transposable elements such as plasmids, integrated and conjugated plasmids within the chromosome and phage insertions. Using MOB-suit to reconstruct plasmids from WGS data, we determined that all these genes were chromosomal and not plasmid-borne. Further using ICEfinder, we determined that these unique genes did not belong to integrative and conjugative elements (ICEs). Using Prophage hunter, we identified that the *hlyE* gene found in the Armenian isolates was located within the active prophage region (Figure 4A) with the closest match to Stx2-converting phage Stx2a_F349 (96% homology 9% coverage of prophage gene region). The *ireA* gene region (Figure 4B) and the *lutABCP/ hsdMSR* operons gene region (Figure 4C) was not predicted by prophage hunter as prophage insertion region. However, based on genes found within these loci (the unique Inovirus Gp2 family protein gene found both in the isolates of this study and in DABGLY01 (recovered in France), adjacent to *ireA* gene and the *lutABC* and *hsdMSR* operons adjacent to the prophage integrase gene *intA* near and tRNA-phe), we hypothesise that a possible horizontal transfer of genes could have occurred in these regions via prophages.

The *ireA* virulent gene was present in other *E. coli* ST127 isolates, but the adjacent *incD* and the Inovirus Gp2 protein unique to the Armenian and DABGLY01 isolates were absent. To investigate further, we conducted a comparative analysis of the *ireA* gene loci of all the *E. coli* ST127 isolates. The main difference in *ireA* gene loci was that all isolates recovered in our study and DABGLY01 harboured two additional insertion sequence transposases instead of a hypothetical protein gene. In addition, the Inovirus gtp2 protein family gene found in the *ireA* locus shared 87.24% protein similarity (Figure 5).

To determine whether the unique genes genomic regions detected in the Armenian (ARM64, ARM66, ARM75 and ARM88) and French DABGLY01 isolates were also present in other *E. coli* isolates from different sequence types, we conducted a sequence similarity search against the entire ENA archive (Tables S5, S6 and S7). We found that each unique genomic region had 100% homology with *E. coli* isolates belonging to other sequence types. In addition, the *ireA* (Table S5) and *hlyE* (Table S6) gene regions had 100% homology with those found in *E. coli* ST761 and ST93 isolates recovered from pig's faeces in Canada and chicken meat in South Korea respectively. For the *lutABC*/*hsdMS* genomic region (Table S7), we found multiple *E. coli* isolates from the ENA archive that had 100% homology with the Armenian ST127 isolates that were sequence types ST1485, ST40, ST117, ST1266, ST95 and ST117, isolated from China, Japan, USA, France, Switzerland and the United Kingdom; and sources including mammalian, avian, human and wastewater. However, we did not find isolates >=99.99% homology containing all three genomic regions; instead, we detected multiple isolates that harboured *ireA* and *hlyA* genomic region and the *ireA* and *lutABC*/*hsdMSR* genomic region. However, all three genome regions were identified in various isolates belonging to ST117 sequence type.

Antibiotic resistance genotype

The antimicrobial resistance genes and mutations of all Armenian *E. coli* ST127 isolates were compared to 168 WGS data of *E. coli* ST127 isolates obtained from the ENA archive. In total, we identified 120 different antimicrobial genes/mutations within the resistome of these isolates (Figure 6). Forty-seven of these genes/mutations were present in all 172 *E. coli* ST127 isolates we analysed. These core antibiotic resistance genes/mutations are known to induce resistance to antibiotics, including those belonging to classes of aminocoumarin, aminoglycoside, carbapenem, cephalosporin, cephamycin, diaminopyrimidine, fluoroquinolone, fosfomycin, glycylcycline, macrolide, monobactam, nitroimidazole, nucleoside, penam, peptide, phenicol, rifamycin and tetracycline.

All four isolates recovered in this study possessed various degrees of antibiotic resistance gene/mutations, including one isolate (ARM64) possessing 59 antibiotic resistance gene/mutations; one isolate (ARM88) 57 antibiotic resistance gene/mutations and two isolates (ARM66 and ARM75) 56 antibiotic resistance gene/mutations. This is more than the mean average of antibiotic resistance genes compared to *E. coli* ST127 isolates obtained from the ENA archive (mean average 55 range 52 to 67 genes) that have been reported previously. In addition to this, a plasmid mediated quinolone resistance protein gene *qnrD1* was unique to the ARM88 isolate recovered in this study.

60 out of 172 (34.88%) of the ST127 isolates, and all 4 Armenian isolates which were recovered in this study, possessed genes that encode for ESBL, including but not limited to CTX (n=17), TEM (n=48) and OXA-1(n=3). The most common CTX gene was $bla_{\text{CTX-M-3}}$ (n=16), also found in three of the Armenian isolates (ARM66, ARM75 and ARM88), $bla_{\text{CTX-M-15}}$ (n=15), $bla_{\text{CTX-M-211}}$ (n=6), $bla_{\text{CTX-M-55}}$ (n=4), $bla_{\text{CTX-M-216}}$ and $bla_{\text{CTX-M-236}}$ (n=2) which was found in one of the Armenian isolate (ARM64) and $bla_{\text{CTX-M-14}}$ (n=2). The most common TEM gene was $bla_{\text{TEM-1}}$ (n=45) also found in ARM64, $bla_{\text{TEM-196}}$ (n=4), $bla_{\text{TEM-112}}$ (n=2), $bla_{\text{TEM-160}}$ (n=1), $bla_{\text{TEM-234}}$ (n=1), $bla_{\text{TEM-34}}$ (n=1) and $bla_{\text{TEM-91}}$ (n=1). In addition, carbapenem-resistance genes $bla_{\text{NMD-1}}$ (n=1) and $bla_{\text{IPM-4}}$ (n=1) were detected in one of each *E. coli* ST127 isolates, however only in those obtained from the ENA archive. We also found the antibiotic resistance gene/mutations, which were found in nearly all *E. coli* ST127 isolates from the ENA archive and in the all Armenian *E. coli* ST127, including multidrug efflux *mdtE*, *mdtF*, the *gadW* and *gadX*; the pmr phosphoethanolamine transferase gene *eptA* (found in 171/172 isolates except for *gadX* which was found in 170/172 of isolates), the elfamycin antibiotic resistance EF-Tu mutation (169/172 of the *E. coli* ST127 isolates) and the tetracycline efflux pump genes *tet(B)* and *tetR* (32/172 of the *E. coli* ST127 isolates). In addition, the aminoglycoside acetyltransferase gene *AAC(3)-IId* was identified in ARM64 (3/172 *E. coli* ST127 isolates).

Plasmid typing

Plasmid replicon was identified in 72% (124/172) *E. coli* ST127 isolates, including in two Armenian isolates (ARM64 and ARM66). 36 different plasmid replicons were identified, of which IncFIB (AP001918) 1 was the most common (n=74), followed by Col156_1 (n=71) and IncFII (29) 1_pUTI89 (n=48) (Figure 7). Most number of plasmid replicons identified was 6 (n=2), followed by 5 (n=6), 4 (n=17), 3 (n=48), 2 (n=20) and 1 (n=31). ARM64 isolates harboured 3 plasmid replicons, including IncFII_1(found in 8.14% of ST127 isolates), IncI1_1_Alpha (found in 5.23% of ST127 isolates) and IncX1 (4.07% of ST127 isolates). Moreover, ARM66 harboured one plasmid replicon, IncI2_1, that was unique to it. Overall, we found that isolates based on the hierarchy cluster heatmap of their plasmid replicons do not group with their BAP cluster groups. We did not detect any known plasmid replicon for the ARM75 and ARM88 isolates despite MOB-suite identifying part of the genome sequence as plasmid associated.

Discussion

Studies investigating the potential of DNA sequencing of pathogens causing infectious diseases are seriously lacking in Armenia, hampering global efforts to track, trace and contain infectious disease outbreaks. In this study, we report for the first-time whole genome analysis of ESBL UPEC ST127 isolates recovered from hospitalised patients in Armenia and compare them with other *E. coli* ST127 retrieved from the ENA (recovered from multiple sources) and identify the mode of horizontal transfer of the unique genes found only in Armenian isolates. Interestingly, all Armenian *E. coli* ST127 isolates

were identified as UPEC ESBL and belonged to the same genetic lineage. Despite these isolates were recovered from patients in five different hospitals, their inferred divergence date of the most recent common ancestor was around 2000 (95% CI date: 1995 to 2006). The close genetic similarities of the Armenian isolates indicated that ST127 was potentially a dominant lineage causing urinary tract infections in Armenia at the time of this study (January to September 2019). In addition, we found that a French UPEC ESBL isolate was the only other isolate from the ENA archive that was part of the same genetic lineage and shared the most recent common ancestor with the Armenian isolates. Furthermore, we identified unique genes near to known prophage genes in both the Armenian and French isolate clusters suggesting that these genes were horizontally transferred via a prophage. These unique genes and the associated genomic regions were novel to the ST127 lineage but were present in other *E. coli* sequence types recovered from different human (including human urine samples) and animal sources. However, none of these isolates was shown to harbour all three genome regions, suggesting a unique combination of genes that are only found in the Armenian and French UPEC ST127 isolate clusters. As there is a large Armenian community in France with relevant family links in Armenia, it seems plausible that the *E. coli* ST127 was transmitted between countries.

The most notable gene found in the Armenian and French ESBL *E. coli* ST127 lineage was the pore forming toxin hemolysin E gene *hlyE* (also known as *clyA* and *sheA*). Hemolysin E has been identified across different species, including *Salmonella typhi* and *Shigella flexneri*, with a haemolytic action towards mammalian cells in anaerobic conditions (51). Moreover, the *hlyE* gene has been identified in APEC, UPEC isolates recovered from companion animals, and ExPEC isolates recovered from humans but has never been described as a virulence factor associated cause of urinary tract infections from *in vivo* studies (4, 52–54, 54, 55). Moreover, we found the haemolysin A gene *hlyA*, which is unrelated to *hlyE* was ubiquities in the Armenian *E. coli* ST127 isolates and has previously been described as an important virulence factor in causing pyelonephritis (51, 56, 57). Therefore, it is plausible that both genes can act synergistically to cause haemolysis in urinary tract infections; however, no such data have been reported up to date due to the rarity of isolates harbouring both *hlyA* and *hlyE* genes (53, 54). The *hlyE* gene region was determined to be a part of a prophage insertion region using Prophage Hunter software (48). We found that the small portion of the prophage region was relatively similar to an Stx2-converting phage Stx2a_F349; a phage well known for horizontal transfer of the Shiga toxin (Stx) in Stx-producing *E. coli* (STEC). However, scientific evidence obtained in laboratory conditions has shown that stx2-phages can also infect ExPEC isolates, although no evidence has ever been speculated of the prophage inserting the Shiga toxin gene into ExPEC isolates (58). Nevertheless, the blast search analysis of these genes in the *hlyE* prophage region identified two genes associated with prophages that were not originated from an Stx2-converting phage. These genes have shown homology to Inovirus-type Gp2 family protein and the transcriptional regulator gene *alpA* gene associated with cryptic P4-like prophage CP4-57. Therefore, we can hypothesise that the prophage insertion region of *hlyE* originated from an unknown prophage species.

Further analysis of other unique gene regions found in the Armenian and French cluster isolates identified one of the unique genes as a prophage gene (unique if they were less than <90% protein homology) or unique genes adjacent to prophage genes in the bacterial chromosome. For example, we identified another gene that had homology to an Inovirus Gp2 family protein unique to our isolates along with IS3 family transposase IS2 near TonB-dependent siderophore receptor gene *ireA* involved in the increased uptake of iron within an iron restricted environment such as human urine (59). The *ireA* was found in five other isolates retrieved from ENA that were recovered from companion animals and human urine samples but showed no direct phylogenetic or geographical linkage to each other. Moreover, we also found that other *ireA* ST127 isolates had an Inovirus Gp2 family protein gene;

however, these genes only shared 87.24% homology at a protein level. The difference in homology may be linked to high mutation rates within filamentous phages due to being single-stranded DNA genome (60). We also found a difference in the genome region of the Armenian and French isolates cluster as they had an insertion sequence IS3 family transposase IS2 elements between *ireA* and Inovirus Gp2 family protein instead of a hypothetical protein found within the other ST127 *ireA* isolates. Moreover, a truncated version of IS255 family transposase ISod4 was not truncated in the other ST127 *ireA* isolates. The insertion sequence acquisition and changes within the Armenian and French isolates cluster may have a role in the immobilisation and stability of the prophage region, which may suggest why we did not find other *E. coli* ST127 clusters containing the *ireA* gene (61, 62).

Furthermore, the Armenian and French isolates cluster had a genomic region that contained 11 unique genes, which were adjacent to a prophage integrase gene *intA* and tRNA-phe gene. Downstream of these genes were the unique genes to the Armenian and French cluster, including a peptide ABC transporter substrate-binding protein, lactate utilisation protein operon *lutABC* (also known as *ykgEFG* in *E. coli*) and the endonuclease restriction modification operon *hsdMSR*. To our knowledge, there have been no reports of involvement of these genes in UPEC growth and survival in the urinary tract. Nonetheless, we can hypothesise that the peptide ABC transporter substrate-binding protein gene, *lutABC* operon, along with *lutP* permease, may have a role in utilising the lactate as an energy source along with other metabolites found in the urine; however, further studies are required (63–65). In addition, the *hsdMSR* operon may have acted initially as a super-infection exclusion mechanism to prevent secondary bacteriophage infection (66). Moreover, *hsdMSR* operon may act as a mechanism in gene regulation within the urinary tract by DNA methylation (67, 68).

Although we were not able to find a complete arsenal of prophage gene integration within all these genomic regions, we hypothesise that the most likely mode of horizontal transfer of these genomic regions was through a prophage and that the prophage genes have decayed over time due to point mutations and deletions to remove unwanted and toxic genomic features, leaving behind remanets of prophage genes *(69)*. Furthermore, based on the date of divergence (1966 date CI: 1956 to 1977); of the UK isolates recovered from blood culture in 2011, which was the closest related isolate to Armenian and French cluster, we cannot determine when these genomic regions integrated into Armenian and French *E. coli* ST127 lineage; whether it had happened before or after diverging from the UK isolates.

Although we did not find antibiotic resistance gene mutations and plasmid replicons that were ubiquities and unique to the Armenia isolates, we did identify that the quinolone resistance protein gene *qnrD1* was unique to ARM88 and the plasmid replicon genes and IncI2_1 were unique to ARM66. The *qnrD1* is generally rare in *E. coli* isolates and is more commonly identified in *Proteus* and *Providencia* species; whereas the IncX3_1 and IncI2_1 are generally associated with the spread of carbapenems and colistin antibiotic resistance genes (70–75). However, we could not link it to any plasmid because we could not identify a known plasmid replicon with an ARM88 plasmid sequence. In addition, we did not detect any antibiotic resistance gene associated with carbapenems and colistin antibiotic resistance in ARM64 and ARM66. In correlation with phenotypic data, we determined that ARM88 *qnrD1* showed known resistance to fluoroquinolone antibiotics (noroflaxin and levofloxacin). Intriguingly, we found 46 antibiotic resistance determents that were ubiquities in all *E. coli* ST127 isolates we studied, which coincides with Goldstone and Smith, 2017 analysis that described 50 antibiotic resistance genes as being part of the core resistome (=>95% of the isolates) and may have a role in providing basal-level resistance towards a diverse number of antimicrobial compounds. However, Goldstone and Smith, 2017 also pointed out that some of these core antibiotic resistance genes may not be involved in antibiotic resistance in *E. coli* due to a lack of evidence in the literature but have been involved in antibiotic

resistance in other bacterial species. Our findings indicate that beta-lactam and cephalosporin antibiotic resistance phenotype in the Armenian isolates is attributed by the ESBL associated genes *bla*$_{\text{CTX-M-3}}$, bla$_{\text{CTX-M-236}}$ and *bla*$_{\text{TEM-1.}}$ However, we were unable to determine the antibiotic resistance genotype that infers phenotypic resistance to aminoglycoside antibiotic amikacin in ARM88 and resistance to fluoroquinolone antibiotics in norfloxacin and levofloxacin in ARM75, intermediate resistance to the carbapenem antibiotics and imipenem resistance in ARM66. Moreover, the antibiotic resistance phenotypes observed in our study were consistent with previous studies reporting that ST127 were sensitive towards a wide range of antibiotics compared to other UPEC sequence types and that aminoglycoside and fluoroquinolones antibiotic resistances were not commonly found in ST127 isolates along with ESBL production.(5, 77, 78).

The main limitation of this study is the small sample size; however, this is the first genomic analysis of ESBL UPEC ST127 isolates recovered from patients in Armenia. These isolates were recovered from different hospitals in different regions of Armenia but belonged to the same genetic lineage and have a recent common shared ancestor. In addition, all isolates shared virulence and metabolic genes acquired via horizontal transfer that were not found in other *E. coli* ST127 isolates except in one phylogenetically closely related UPEC isolate from France. Further whole genome sequencing surveillance is necessary to better understand the molecular epidemiology of the Armenian isolates, the role of some of these genes and to determine whether ST127 is the dominate lineage causing urinary tract infections in Armenia. Such surveillance studies will contribute to global efforts to tackle antibiotic resistance as bacteria carrying AMR genes do not recognise borders.

Conflict of Interest

The authors declare no potential conflicts of interest.

Funding

This work was supported by the QR Global Challenges Research Fund.

Reference

1. Flores-Mireles AL, Walker JN, Caparon M, Hultgren SJ. 2015. Urinary tract infections: epidemiology, mechanisms of infection and treatment options. Nat Rev Microbiol 13:269–284.

2. Kot B. 2019. Antibiotic Resistance Among Uropathogenic Escherichia coli. Pol J Microbiol 68:403–415.

3. Church LWP. 2008. Enterobacteriaceae, p. 945–952. *In* Schlossberg, D (ed.), Clinical Infectious Disease. Cambridge University Press, Cambridge.

4. Sarowska J, Futoma-Koloch B, Jama-Kmiecik A, Frej-Madrzak M, Ksiazczyk M, Bugla-Ploskonska G, Choroszy-Krol I. 2019. Virulence factors, prevalence and

potential transmission of extraintestinal pathogenic Escherichia coli isolated from different sources: recent reports. Gut Pathog 11:10.

5. Gibreel TM, Dodgson AR, Cheesbrough J, Fox AJ, Bolton FJ, Upton M. 2012. Population structure, virulence potential and antibiotic susceptibility of uropathogenic Escherichia coli from Northwest England. Journal of Antimicrobial Chemotherapy 67:346–356.

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

7. Paramita RI, Nelwan EJ, Fadilah F, Renesteen E, Puspandari N, Erlina L. 2020. Genome-based characterization of Escherichia coli causing bloodstream infection through next-generation sequencing. PLoS One 15:e0244358.

8. Desvaux M, Dalmasso G, Beyrouthy R, Barnich N, Delmas J, Bonnet R. 2020. Pathogenicity Factors of Genomic Islands in Intestinal and Extraintestinal Escherichia coli. Front Microbiol 11:2065.

9. Messerer M, Fischer W, Schubert S. 2017. Investigation of horizontal gene transfer of pathogenicity islands in Escherichia coli using next-generation sequencing. PLOS ONE 12:e0179880.

10. von Wintersdorff CJH, Penders J, van Niekerk JM, Mills ND, Majumder S, van Alphen LB, Savelkoul PHM, Wolffs PFG. 2016. Dissemination of Antimicrobial Resistance in Microbial Ecosystems through Horizontal Gene Transfer. Frontiers in Microbiology 7.

11. Juhas M, van der Meer JR, Gaillard M, Harding RM, Hood DW, Crook DW. 2009. Genomic islands: tools of bacterial horizontal gene transfer and evolution. FEMS Microbiol Rev 33:376–393.

12. Rossolini GM, D'Andrea MM, Mugnaioli C. 2008. The spread of CTX-M-type extended-spectrum β-lactamases. Clinical Microbiology and Infection 14:33–41.

13. Bezabih YM, Bezabih A, Dion M, Batard E, Teka S, Obole A, Dessalegn N, Enyew A, Roujeinikova A, Alamneh E, Mirkazemi C, Peterson GM, Bezabhe WM. 2022. Comparison of the global prevalence and trend of human intestinal carriage of ESBL-producing Escherichia coli between healthcare and community settings: a systematic review and meta-analysis. JAC-Antimicrobial Resistance 4:dlac048.

14. Devi LS, Broor S, Rautela RS, Grover SS, Chakravarti A, Chattopadhya D. 2020. Increasing Prevalence of Escherichia coli and Klebsiella pneumoniae Producing CTX-M-Type Extended-Spectrum Beta-Lactamase, Carbapenemase, and NDM-1 in Patients from a Rural Community with Community Acquired Infections: A 3-Year Study. Int J Appl Basic Med Res 10:156–163.

15. Martischang R, François P, Cherkaoui A, Gaïa N, Renzi G, Agostinho A, Perez M, Graf CE, Harbarth S. 2021. Epidemiology of ESBL-producing Escherichia coli from repeated prevalence studies over 11 years in a long-term-care facility. Antimicrobial Resistance & Infection Control 10:148.

16. Elankumaran P, Browning GF, Marenda MS, Reid CJ, Djordjevic SP. 2022. Close genetic linkage between human and companion animal extraintestinal pathogenic Escherichia coli ST127. Current Research in Microbial Sciences 3:100106.

17. Halaji M, Shahidi S, Ataei B, Atapour A, Feizi A, Havaei SA. 2021. Molecular epidemiology of blaCTX-M gene-producing uropathogenic Escherichia coli among Iranian kidney transplant patients: clonal dissemination of CC131 and CC10. Ann Clin Microbiol Antimicrob 20:65.

18. Alba P, Leekitcharoenphon P, Franco A, Feltrin F, Ianzano A, Caprioli A, Stravino F, Hendriksen RS, Bortolaia V, Battisti A. 2018. Molecular Epidemiology of mcr-Encoded Colistin Resistance in Enterobacteriaceae From Food-Producing Animals in Italy Revealed Through the EU Harmonized Antimicrobial Resistance Monitoring. Frontiers in Microbiology 9.

19. Johnson JR, Russo TA. 2018. Molecular Epidemiology of Extraintestinal Pathogenic Escherichia coli. EcoSal Plus 8.

20. Takawira FT, Pitout JD, Thilliez G, Mashe T, Gutierrez AV, Kingsley RA, Peirano G, Matheu J, Midzi SM, Mwamakamba LW, Gally DL, Tarupiwa A, Mukavhi L, Ehlers MM, Mtapuri-Zinyowera S, Kock MM. 2021. Molecular epidemiology of extended-spectrum beta-lactamase-producing extra-intestinal pathogenic Escherichia coli strains over a 2-year period (2017-2019) from Zimbabwe. EJCMID https://doi.org/10.1007/s10096-021-04379-z.

21. Iskandar K, Molinier L, Hallit S, Sartelli M, Hardcastle TC, Haque M, Lugova H, Dhingra S, Sharma P, Islam S, Mohammed I, Naina Mohamed I, Hanna PA, Hajj SE, Jamaluddin NAH, Salameh P, Roques C. 2021. Surveillance of antimicrobial resistance in low- and middle-income countries: a scattered picture. Antimicrobial Resistance & Infection Control 10:63.

22. Shen H, Ma H. 2013. Molecular Epidemiology in Low- and Middle-Income Countries, p. 157–173. *In* Soliman, A, Schottenfeld, D, Boffetta, P (eds.), Cancer Epidemiology. Oxford University Press.

23. Data for Armenia, Upper middle income | Data. World Bank. https://data.worldbank.org/?locations=AM-XT. Retrieved 18 May 2022.

24. Mkrtchyan HV, Xu Z, Yacoub M, Ter-Stepanyan MM, Karapetyan HD, Kearns AM, Cutler RR, Pichon B, Hambardzumyan AD. 2017. Detection of diverse genotypes of Methicillin-resistant Staphylococcus aureus from hospital personnel and the environment in Armenia. Antimicrob Resist Infect Control 6:19.

25. Jamrozy D, Misra R, Xu Z, Ter-Stepanyan MM, Kocharyan KS, Cave R, Hambardzumyan AD, Mkrtchyan HV. 2019. Novel Methicillin-Resistant Staphylococcus aureus CC8 Clone Identified in a Hospital Setting in Armenia. Frontiers in Microbiology 10.

26. Mkrtchyan HV, Russell CA, Wang N, Cutler RR. 2013. Could Public Restrooms Be an Environment for Bacterial Resistomes? PLOS ONE 8:e54223.

27. "The European Committee on Antimicrobial Susceptibility Testing. Breakpoint tables for interpretation of MICs and zone diameters. Version 11.0, 2021. http://www.eucast.org.

28. Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/. Retrieved 16 November 2021.

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

30. Treangen T, Ondov BD, Koren S, Phillippy AM. 2014. The Harvest suite for rapid core-genome alignment and visualization of thousands of intraspecific microbial genomes. Genome biology 15:524.

31. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. Molecular Biology and Evolution 32:268–274.

32. Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. 2018. UFBoot2: Improving the Ultrafast Bootstrap Approximation. Molecular Biology and Evolution 35:518–522.

33. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. 6. Nat Methods 14:587–589.

34. Letunic I, Bork P. 2021. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Research 49:W293–W296.

35. Suchard MA, Lemey P, Baele G, Ayres DL, Drummond AJ, Rambaut A. 2018. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. Virus Evolution 4:vey016.

36. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30:2068–2069.

587 37. Carattoli A, Zankari E, García-Fernández A, Voldby Larsen M, Lund O, Villa L, Møller
588 Aarestrup F, Hasman H. 2014. In Silico Detection and Typing of Plasmids using
589 PlasmidFinder and Plasmid Multilocus Sequence Typing. Antimicrob Agents
590 Chemother 58:3895–3903.

591 38. Chen L, Yang J, Yu J, Yao Z, Sun L, Shen Y, Jin Q. 2005. VFDB: a reference database
592 for bacterial virulence factors. Nucleic Acids Res 33:D325-328.

593 39. Joensen KG, Scheutz F, Lund O, Hasman H, Kaas RS, Nielsen EM, Aarestrup FM.
594 2014. Real-time whole-genome sequencing for routine typing, surveillance, and
595 outbreak detection of verotoxigenic Escherichia coli. J Clin Microbiol 52:1501–1510.

596 40. Alcock BP, Raphenya AR, Lau TTY, Tsang KK, Bouchard M, Edalatmand A, Huynh
597 W, Nguyen A-LV, Cheng AA, Liu S, Min SY, Miroshnichenko A, Tran H-K, Werfalli
598 RE, Nasir JA, Oloni M, Speicher DJ, Florescu A, Singh B, Faltyn M, Hernandez-
599 Koutoucheva A, Sharma AN, Bordeleau E, Pawlowski AC, Zubyk HL, Dooley D,
600 Griffiths E, Maguire F, Winsor GL, Beiko RG, Brinkman FSL, Hsiao WWL, Domselaar
601 GV, McArthur AG. 2020. CARD 2020: antibiotic resistome surveillance with the
602 comprehensive antibiotic resistance database. Nucleic Acids Res 48:D517–D525.

603 41. Bessonov K, Laing C, Robertson J, Yong I, Ziebell K, Gannon VPJ, Nichani A, Arya G,
604 Nash JHE, Christianson S. 2021. ECTyper: in silico Escherichia coli serotype and
605 species prediction from raw and assembled whole-genome sequence data. Microb
606 Genom 7.

607 42. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MTG, Fookes M, Falush
608 D, Keane JA, Parkhill J. 2015. Roary: rapid large-scale prokaryote pan genome analysis.
609 Bioinformatics 31:3691–3693.

610 43. Brynildsrud O, Bohlin J, Scheffer L, Eldholm V. 2016. Rapid scoring of genes in
611 microbial pan-genome-wide association studies with Scoary. Genome Biology 17:238.

612 44. Katoh K, Standley DM. 2013. MAFFT Multiple Sequence Alignment Software Version
613 7: Improvements in Performance and Usability. Mol Biol Evol 30:772–780.

614 45. Darling ACE, Mau B, Blattner FR, Perna NT. 2004. Mauve: Multiple Alignment of
615 Conserved Genomic Sequence With Rearrangements. Genome Res 14:1394–1403.

616 46. Robertson J, Nash JHE. 2018. MOB-suite: software tools for clustering, reconstruction
617 and typing of plasmids from draft assemblies. Microb Genom 4.

618 47. Liu M, Li X, Xie Y, Bi D, Sun J, Li J, Tai C, Deng Z, Ou H-Y. 2019. ICEberg 2.0: an
619 updated database of bacterial integrative and conjugative elements. Nucleic Acids
620 Research 47:D660–D665.

621 48. Song W, Sun H-X, Zhang C, Cheng L, Peng Y, Deng Z, Wang D, Wang Y, Hu M, Liu
622 W, Yang H, Shen Y, Li J, You L, Xiao M. 2019. Prophage Hunter: an integrative
623 hunting tool for active prophages. Nucleic Acids Research 47:W74–W80.

624 49. Madeira F, Park YM, Lee J, Buso N, Gur T, Madhusoodanan N, Basutkar P, Tivey
625 ARN, Potter SC, Finn RD, Lopez R. 2019. The EMBL-EBI search and sequence
626 analysis tools APIs in 2019. Nucleic Acids Res 47:W636–W641.

627 50. Tonkin-Hill G, Lees JA, Bentley SD, Frost SDW, Corander J. 2018. RhierBAPS: An R
628 implementation of the population clustering algorithm hierBAPS. Wellcome Open Res
629 3:93.

630 51. Hunt S, Green J, Artymiuk PJ. 2010. Hemolysin E (HlyE, ClyA, SheA) and Related
631 Toxins, p. 116–126. *In* Anderluh, G, Lakey, J (eds.), Proteins Membrane Binding and
632 Pore Formation. Springer, New York, NY.

633 52. Hutton TA, Innes GK, Harel J, Garneau P, Cucchiara A, Schifferli DM, Rankin SC.
634 2018. Phylogroup and virulence gene association with clinical characteristics of
635 Escherichia coli urinary tract infections from dogs and cats. J VET Diagn Invest 30:64–
636 70.

637 53. Kerényi M, Allison HE, Bátai I, Sonnevend Á, Emödy L, Plaveczky N, Pál T. 2005.
638 Occurrence of hlyA and sheA Genes in Extraintestinal Escherichia coli Strains. Journal
639 of Clinical Microbiology https://doi.org/10.1128/JCM.43.6.2965-2968.2005.

640 54. Lorenz SC, Son I, Maounounen-Laasri A, Lin A, Fischer M, Kase JA. 2013. Prevalence
641 of Hemolysin Genes and Comparison of ehxA Subtype Patterns in Shiga Toxin-
642 Producing Escherichia coli (STEC) and Non-STEC Strains from Clinical, Food, and
643 Animal Sources. Applied and Environmental Microbiology
644 https://doi.org/10.1128/AEM.02200-13.

645 55. Massella E, Giacometti F, Bonilauri P, Reid CJ, Djordjevic SP, Merialdi G, Bacci C,
646 Fiorentini L, Massi P, Bardasi L, Rubini S, Savini F, Serraino A, Piva S. 2021.
647 Antimicrobial Resistance Profile and ExPEC Virulence Potential in Commensal
648 Escherichia coli of Multiple Sources. 4. Antibiotics 10:351.

649 56. Ristow LC, Welch RA. 2016. Hemolysin of uropathogenic Escherichia coli: A cloak or
650 a dagger? Biochimica et Biophysica Acta (BBA) - Biomembranes 1858:538–545.

651 57. Wang C, Li Q, Lv J, Sun X, Cao Y, Yu K, Miao C, Zhang Z-S, Yao Z, Wang Q. 2020.
652 Alpha-hemolysin of uropathogenic Escherichia coli induces GM-CSF-mediated acute
653 kidney injury. 1. Mucosal Immunol 13:22–33.

654 58. Tozzoli R, Grande L, Michelacci V, Ranieri P, Maugliani A, Caprioli A, Morabito S.
655 2014. Shiga toxin-converting phages and the emergence of new pathogenic Escherichia
656 coli: a world in motion. Front Cell Infect Microbiol 4:80.

657 59. Torres-Puig S, García V, Stærk K, Andersen TE, Møller-Jensen J, Olsen JE, Herrero-
658 Fresno A. 2022. “Omics” Technologies - What Have They Told Us About
659 Uropathogenic Escherichia coli Fitness and Virulence During Urinary Tract Infection?
660 Frontiers in Cellular and Infection Microbiology 12.

661 60. Peck KM, Lauring AS. 2018. Complexities of Viral Mutation Rates. Journal of Virology
662 https://doi.org/10.1128/JVI.01031-17.

663 61. Ooka T, Ogura Y, Asadulghani M, Ohnishi M, Nakayama K, Terajima J, Watanabe H,
664 Hayashi T. 2009. Inference of the impact of insertion sequence (IS) elements on
665 bacterial genome diversification through analysis of small-size structural
666 polymorphisms in Escherichia coli O157 genomes. Genome Res 19:1809–1816.

667 62. Ramisetty BCM, Sudhakari PA. 2019. Bacterial ‘Grounded’ Prophages: Hotspots for
668 Genetic Renovation and Innovation. Frontiers in Genetics 10.

669 63. Hua Q, Joyce AR, Palsson BØ, Fong SS. 2007. Metabolic Characterization of
670 Escherichia coli Strains Adapted to Growth on Lactate. Appl Environ Microbiol
671 73:4639–4647.

64. Mann R, Mediati DG, Duggin IG, Harry EJ, Bottomley AL. 2017. Metabolic Adaptations of Uropathogenic E. coli in the Urinary Tract. Frontiers in Cellular and Infection Microbiology 7.

65. Núñez MF, Kwon O, Wilson TH, Aguilar J, Baldoma L, Lin ECC. 2002. Transport of L-Lactate, D-Lactate, and glycolate by the LldP and GlcA membrane carriers of Escherichia coli. Biochem Biophys Res Commun 290:824–829.

66. Dedrick RM, Jacobs-Sera D, Guerrero Bustamante CA, Garlena RA, Mavrich TN, Pope WH, Reyes JCC, Russell DA, Adair T, Alvey R, Bonilla JA, Bricker JS, Brown BR, Byrnes D, Cresawn SG, Davis WB, Dickson LA, Edgington NP, Findley AM, Golebiewska U, Grose JH, Hayes CF, Hughes LE, Hutchison KW, Isern S, Johnson AA, Kenna MA, Klyczek KK, Mageeney CM, Michael SF, Molloy SD, Montgomery MT, Neitzel J, Page ST, Pizzorno MC, Poxleitner MK, Rinehart CA, Robinson CJ, Rubin MR, Teyim JN, Vazquez E, Ware VC, Washington J, Hatfull GF. 2017. Prophage-mediated defense against viral attack and viral counter-defense. Nat Microbiol 2:16251.

67. Casadesús J, Low D. 2006. Epigenetic Gene Regulation in the Bacterial World. Microbiol Mol Biol Rev 70:830–856.

68. Murray NE. 2000. Type I Restriction Systems: Sophisticated Molecular Machines (a Legacy of Bertani and Weigle). Microbiol Mol Biol Rev 64:412–434.

69. Bobay L-M, Touchon M, Rocha EPC. 2014. Pervasive domestication of defective prophages by bacteria. Proceedings of the National Academy of Sciences 111:12127–12132.

70. Albornoz E, Lucero C, Romero G, Quiroga MP, Rapoport M, Guerriero L, Andres P, Rodriguez C, WHONET-Argentina Group, Galas M, Centrón D, Corso A, Petroni A. 2017. Prevalence of Plasmid-Mediated Quinolone Resistance Genes in Clinical Enterobacteria from Argentina. Microbial Drug Resistance 23:177–187.

71. Liu Z, Xiao X, Li Y, Liu Y, Li R, Wang Z. 2019. Emergence of IncX3 Plasmid-Harboring blaNDM–5 Dominated by Escherichia coli ST48 in a Goose Farm in Jiangsu, China. Frontiers in Microbiology 10.

72. Mao Y-C, Chang C-L, Huang Y-C, Su L-H, Lee C-T. 2018. Laboratory investigation of a suspected outbreak caused by Providencia stuartii with intermediate resistance to imipenem at a long-term care facility. J Microbiol Immunol Infect 51:214–219.

73. Meinersmann RJ. 2019. The biology of IncI2 plasmids shown by whole-plasmid multi-locus sequence typing. Plasmid 106:102444.

74. Mouftah SF, Pál T, Darwish D, Ghazawi A, Villa L, Carattoli A, Sonnevend Á. 2019. Epidemic IncX3 plasmids spreading carbapenemase genes in the United Arab Emirates and worldwide. Infect Drug Resist 12:1729–1742.

75. Zhu W, Wang X, Qin J, Liang W, Shen Z. 2020. Dissemination and Stability of the *bla* $_{NDM-5}$ -Carrying IncX3-Type Plasmid among Multiclonal Klebsiella pneumoniae Isolates. mSphere 5:e00917-20.

76. Goldstone RJ, Smith DGE. 2017. A population genomics approach to exploiting the accessory “resistome” of Escherichia coli. Microb Genom 3:e000108.

77. Darling AE, McKinnon J, Worden P, Santos J, Charles IG, Chowdhury PR, Djordjevic SP. 2014. A draft genome of Escherichia coli sequence type 127 strain 2009-46. Gut Pathog 6:32.

78. Fibke CD, Croxen MA, Geum HM, Glass M, Wong E, Avery BP, Daignault D, Mulvey MR, Reid-Smith RJ, Parmley EJ, Portt A, Boerlin P, Manges AR. 2019. Genomic Epidemiology of Major Extraintestinal Pathogenic Escherichia coli Lineages Causing Urinary Tract Infections in Young Women Across Canada. Open Forum Infect Dis 6:ofz431.

Tables and Figures

Table 1: Antibiotic susceptibility profiles of *E. coli* ST127 isolates recovered from urine specimen.

Table 2: Isolation source, country, and BAP cluster of *E. coli* ST127 used in phylogenetic analysis.

Table 3: Unique genes (with known functions) detected in ARM64 and ARM66, ARM75 and ARM88 and DABGLY01

Figure 1: ST127 core maximum likelihood phylogenetic tree and virulence gene heatmap A) phylogenetic tree and virulence gene heatmap, B) correlation matrix of virulent genes and their association to different sources and clusters. Red circle in phylogenetic tree represents Armenian isolates.

Figure 2: MCC time-calibrated phylogenetic tree of ST127 *E. coli*. Blue bars on nodes represent date of divergence 95% CI. A) Full dated MCC dated tree B) Partial : MCC time-calibrated phylogenetic tree showing most recent divergence.

Figure 3: Hierarchy cluster heatmap of accessory genome of ST127 *E. coli* isolates.

Figure 4: Genome loci of unique genes found within Armenian isolates ARM64, ARM66, ARM75, ARM88 and French isolate DABGLY01 A) *hlyE* loci B) *ireA* loci C) *lutABC/hsdMSR* operon loci. Red star unique genes to these isolates, green triangle prophage genes, red block prophage region identified by prophage hunter.

Figure 5: Mauve output of genome loci comparison of *ireA gene.* Red block indicate genome similarities between the two loci.

Figure 6: Hierarchy cluster heatmap of ST127 *E. coli* antibiotic resistant gene/mutation.

753 **Figure 7**: Hierarchy cluster heatmap of ST127 *E. coli* plasmid replicon.
754
755