



UWL REPOSITORY

repository.uwl.ac.uk

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

Cave, Rory, Ter-Stepanyan, Mary M, Kotsinyan, Nune and Mkrtchyan, Hermine (2022) An emerging lineage of uropathogenic extended spectrum β -lactamase Escherichia coli ST127. Microbiology Spectrum, 10 (6).

<http://dx.doi.org/10.1128/spectrum.02511-22>

This is the Published Version of the final output.

UWL repository link: <https://repository.uwl.ac.uk/id/eprint/9605/>

Alternative formats: If you require this document in an alternative format, please contact: open.research@uwl.ac.uk

Copyright: Creative Commons: Attribution 4.0

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy: If you believe that this document breaches copyright, please contact us at open.research@uwl.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



An Emerging Lineage of Uropathogenic Extended Spectrum β -Lactamase *Escherichia coli* ST127

Rory Cave,^a Mary M. Ter-Stepanyan,^{b,c} Nune Kotsinyan,^d  Hermine V. Mkrtchyan^a

^aSchool of Biomedical Sciences, University of West London, London, United Kingdom

^bYerevan State Medical University after Mkhitar Heratsi, Faculty of Public Health, Department of Epidemiology, Yerevan, Republic of Armenia

^cResearch Center of Maternal and Child Health Protection, Yerevan, Armenia

^dNational Centre for Disease Control and Prevention, Yerevan, Armenia

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 β -lactamase (ESBL), UPEC sequence type (ST) 127 isolates that were recovered from patients in five hospitals in Armenia from January to August of 2019. A phylogenetic comparison revealed that our isolates were closely related to each other by their core and accessory genomes, despite having been 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 utilization operon *lutABC*, and endonuclease restriction modification operon *hsdMSR*) were identified in three separate genomic regions that were 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 harbored antibiotic resistance genes, including the ESBL genes *bla*_{CTX-M-3} ($n = 3$), *bla*_{CTX-M-236} ($n = 1$), and *bla*_{TEM-1} ($n = 1$), in addition to a quinolone resistance protein gene *qnrD1* ($n = 1$), which was absent in the ST127 isolates obtained from the 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 the whole-genome sequencing and analysis of ESBL UPEC ST127 isolates recovered from hospitalized 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 that were horizontally acquired in the clusters of Armenian and French isolates that were absent in other ST127 isolates obtained from the ENA. Our findings highlight a possible cross-country transmission between Armenia and France and the idea that the implementation of WGS surveillance could contribute to global efforts in tackling antibiotic resistance, as bacteria carrying antimicrobial resistance (AMR) genes do not recognize borders.

KEYWORDS ESBL, extraintestinal pathogenic *E. coli* (ExPEC), ST127, UTI infections, uropathogenic *Escherichia coli* (UPEC), WGS

Escherichia coli is one of the most common causes of urinary tract infections (UTI), and it is estimated to affect about 150 million people, globally (1). New, emerging lineages present a major challenge for health care settings due to their increased resistance to multiple antibiotics

Editor Bonnie Chase Prokesch, University of Texas Southwestern Medical Center

Copyright © 2022 Cave et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Hermine V. Mkrtchyan, Hermine.mkrtchyan@uwl.ac.uk.

The authors declare no conflict of interest.

Received 1 July 2022

Accepted 8 November 2022

Published 23 November 2022

TABLE 1 Antibiotic susceptibility profiles of *E. coli* ST127 isolates recovered from urine specimen^a

ST	ID	AMP	TZP	AMC	CAZ	CPM	NOR	LVX	AMK	IPM	MEM	CHL
127	ARM64	R	I	R	R	R	S	S	S	S	S	S
127	ARM66	R	I	S	R	R	S	S	S	I	S	S
127	ARM75	R	S	R	R	R	R	R	S	S	S	S
127	ARM88	R	S	R	R	R	S	S	R	S	S	S

^aAMP, ampicillin; TZP, piperacillin-tazobactam; AMC, amoxicillin-clavulanic acid; CAZ, ceftazidime; CPM, cefepime; NOR, norfloxacin; LVX, levofloxacin; AMK, amikacin; IMP, imipenem; MEM, meropenem; CHL, clindamycin.

(2), which contributes to a large number of hospitalizations and associated costs. The uropathogenic *E. coli* (UPEC) pathotype is the primary cause of urinary tract infections and is a part of the broader pathotype designated extraintestinal pathogenic *E. coli* (ExPEC), which is known to cause infections in the bloodstream and in other nonintestinal 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, which is known for its high virulence and low antibiotic resistance potential, compared to other pandemic ExPEC sequence types (5–7). The acquisition of many virulence and antibiotic resistance determinants by UPEC isolates occurs through the 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 evolutionary adaption with little fitness cost under nonselective conditions (11). An example of this is the plasmid-borne CTX-M-type genes that encode extended-spectrum β -lactamases (ESBL), which give the bacteria resistance toward extended-spectrum cephalosporins and monobactams (12). Over the past decade, the prevalence of ESBL-producing *E. coli* has increased globally (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 abilities of whole-genome sequencing (WGS) analyses to trace the sources and transmission routes of infections as well as to identify genes involved in virulence and antibiotic resistance for the implementation of improved infection control practices (6, 18–20). However, studies reporting WGS are limited to some regions of the world (i.e., the United Kingdom, the European Union, and North America), whereas studies from low-middle-income countries are fragmented (21, 22). There have been limited studies utilizing 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 hospitalized patients in Armenia using WGS analysis, 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.

RESULTS

Isolates and antibiotic susceptibility testing. 4 out of the 12 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 the antibiotics tested. All four isolates were resistant to the cephalosporin antibiotics cefepime and ceftazidime as well as the aminopenicillins antibiotic ampicillin. In addition, 3 isolates (ARM64, ARM75, and ARM88) were resistant to the β -lactam antibiotic amoxicillin-clavulanic acid, and 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 and the fluoroquinolone antibiotics norfloxacin and levofloxacin (ARM75). One isolate had an intermediate resistance to the carbapenem antibiotic imipenem (ARM66). All four isolates were sensitive to the β -lactam antibiotic meropenem and were also sensitive to chloramphenicol.

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

Country	Source	BAP
Sri Lanka (<i>n</i> = 1)	Human (<i>n</i> = 137)	BAP1 (<i>n</i> = 7)
Armenia (<i>n</i> = 4)	Cattle (<i>n</i> = 1)	BAP2 (<i>n</i> = 111)
USA (<i>n</i> = 44)	Gazelle (<i>n</i> = 0)	BAP3 (<i>n</i> = 43)
Nigeria (<i>n</i> = 2)	European Mink (<i>n</i> = 1)	BAP4 (<i>n</i> = 9)
United Kingdom (<i>n</i> = 55)	Avian (<i>n</i> = 10)	BAP5 (<i>n</i> = 2)
Canada (<i>n</i> = 34)	Dog (<i>n</i> = 14)	
Japan (<i>n</i> = 1)	Marmoset (<i>n</i> = 2)	
Australia (<i>n</i> = 11)	Striped skunk (<i>n</i> = 2)	
France (<i>n</i> = 2)	Giant panda (<i>n</i> = 1)	
Thailand (<i>n</i> = 1)	Primate (<i>n</i> = 1)	
New Zealand (<i>n</i> = 1)	Cat (<i>n</i> = 1)	
Singapore (<i>n</i> = 4)		
Malaysia (<i>n</i> = 2)		
China (<i>n</i> = 1)		
Denmark (<i>n</i> = 8)		
Germany (<i>n</i> = 1)		

Phylogenetic analysis of *E. coli* ST127. The short reads of four of the *E. coli* ST127 in this study were aligned against 168 *E. coli* ST127 genomes that were obtained from the ENA archive and were previously recovered from 15 countries and 11 different animal sources and belonged to 3 different serotypes (O6:H31 [*n* = 161], O75:H31 [*n* = 3], and -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 were only found from one particular country or source (Fig. 1A). All of the O75:H31 serotype isolates clustered together in clade B. Further detailed analyses of the clusters (using heirBAPS) revealed that the ST127 isolates could be further grouped into 5 distinct clusters. The *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 of the Armenian ST127 isolates were phylogenetically closely related to a French isolate (DABGLY01) that was recovered from a patient (urine sample) in 2015.

A maximum clade credibility (MCC) time-calibrated phylogeny tree of 171 *E. coli* ST127 isolates (excluding the EC536 isolate due to the 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 (Fig. 2). The ST127 had a rate estimate of 9.18×10^{-5} substitutions per site per year and an 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 that was shared by both the Armenian isolates and the French isolate 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 that we analyzed, of which 41 were ubiquities (Fig. 1B). The Armenian isolates harbored 96 (ARM64), 95 (ARM75), 94 (ARM66), and 92 (ARM88) virulence genes, and these numbers are less than the average number (average: 100 genes; range: 63 to 109 genes) of virulence genes, compared to the other ST127 isolates that were analyzed.

All four of the isolates in this study and a phylogenetically closely related French isolate DABGLY01 (<25 SNP difference between the Armenian isolates and the French isolate within the core genomes) harbored the hemolysin E gene (*hlyE*), which was not present in any of the other ST127 isolates. In addition, they also carried the 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 fecal swabs (DABDVM01, DABMUI01, DABXBZ01). However, these isolates did not show any phylogenetic or geographical relationship to each other.

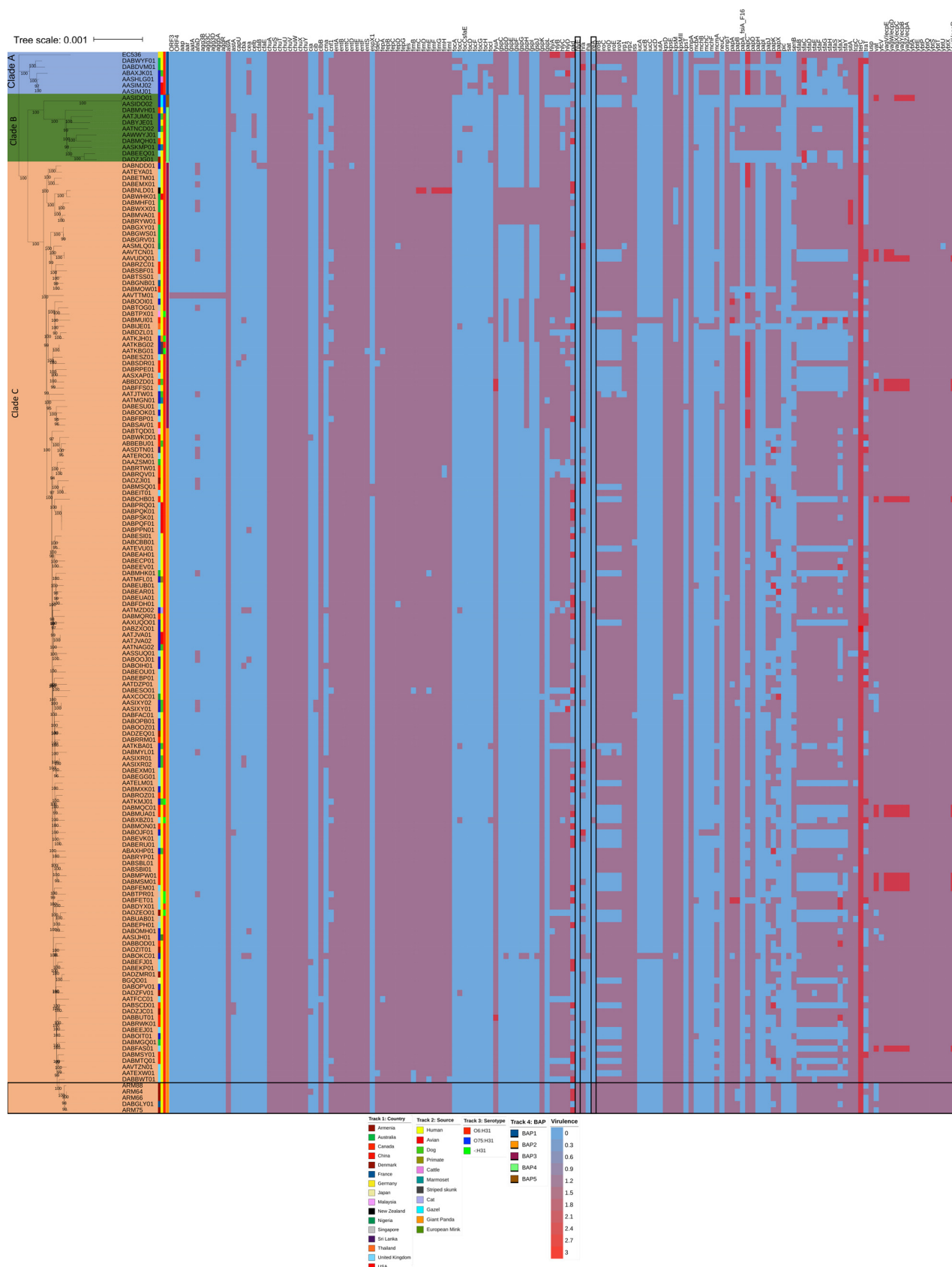


FIG 1 ST127 core maximum likelihood phylogenetic tree and virulence gene heat map. (A) Phylogenetic tree and virulence gene heat map. (B) Correlation matrix of virulent genes and their associations to different sources and clusters. The red circle in the phylogenetic tree represents the Armenian isolates.

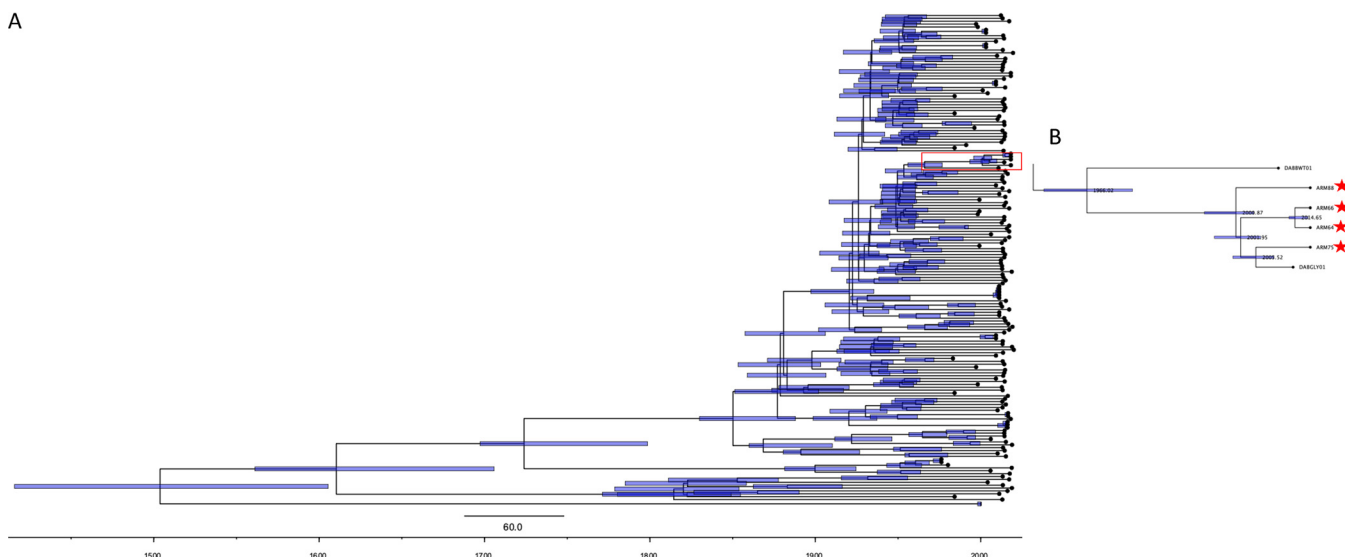


FIG 2 MCC time-calibrated phylogenetic tree of ST127 *E. coli*. The blue bars on the nodes represent the date of divergence 95% confidence intervals. (A) Full dated MCC tree. (B) Partial MCC time-calibrated phylogenetic tree, showing the most recent divergence.

Unique genes and origin of horizontal gene transfer. To further determine whether any of the genes were unique to the Armenian isolates or to the French isolate obtained from the ENA archive, we conducted a pangenome analysis of all of the *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 and 11,191 belonged to the accessory genome. The accessory genome hierarchy clustering heat map (Fig. 3) showed that the Armenian isolates shared many accessory genes with DABGLY01.

Using Scoary, we identified 16 genes that were unique to all of the *E. coli* ST127 Armenian isolates and also to DABGLY01, which was recovered in France. 11 out of the 16 genes identified had a known function (Table 3), including the hemolysin E toxin (*hlyE*), lactate utilization protein operon (*lutABC*), peptide ABC transporter substrate-binding protein, transposase *insD* for insertion element IS2, Inovirus Gp2 family protein, and 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, including the *lutABC* operon gene and the *oppA* gene. In addition, two genes that were located near the *ireA* gene and were unique to the Armenian and DABGLY01 isolates are *InsD* and the Inovirus Gp2 family protein gene. Moreover, we found one additional uncharacterized gene that was unique to the Armenian isolates adjacent to the *hlyE* virulent gene.

To determine the origin of the horizontal gene transfer of these unique genes found in the Armenian isolates, we looked for insertion via transposable elements, such as integrated and conjugated plasmids within the chromosome and phage insertions. Using the MOB-suite software package to reconstruct plasmids from WGS data, we determined that all of these genes were chromosomal and were 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 (Fig. 4A), with the closest match being to Stx2-converting phage Stx2a_F349 (96% homology and 9% coverage of the prophage gene region). The *ireA* gene region (Fig. 4B) and the *lutABC*/*hsdMSR* operons gene region (Fig. 4C) were not predicted to be prophage insertion regions by Prophage Hunter. However, based on genes found within these loci (the unique Inovirus Gp2 family protein gene, which was found both in the isolates of this study and in DABGLY01 [recovered in France] and was adjacent to the *ireA* gene and to the *lutABC* and *hsdMSR* operons that are adjacent to the prophage integrase gene *intA* and near tRNA-phe), we hypothesize that a possible horizontal transfer of genes could have occurred in these regions via prophages.

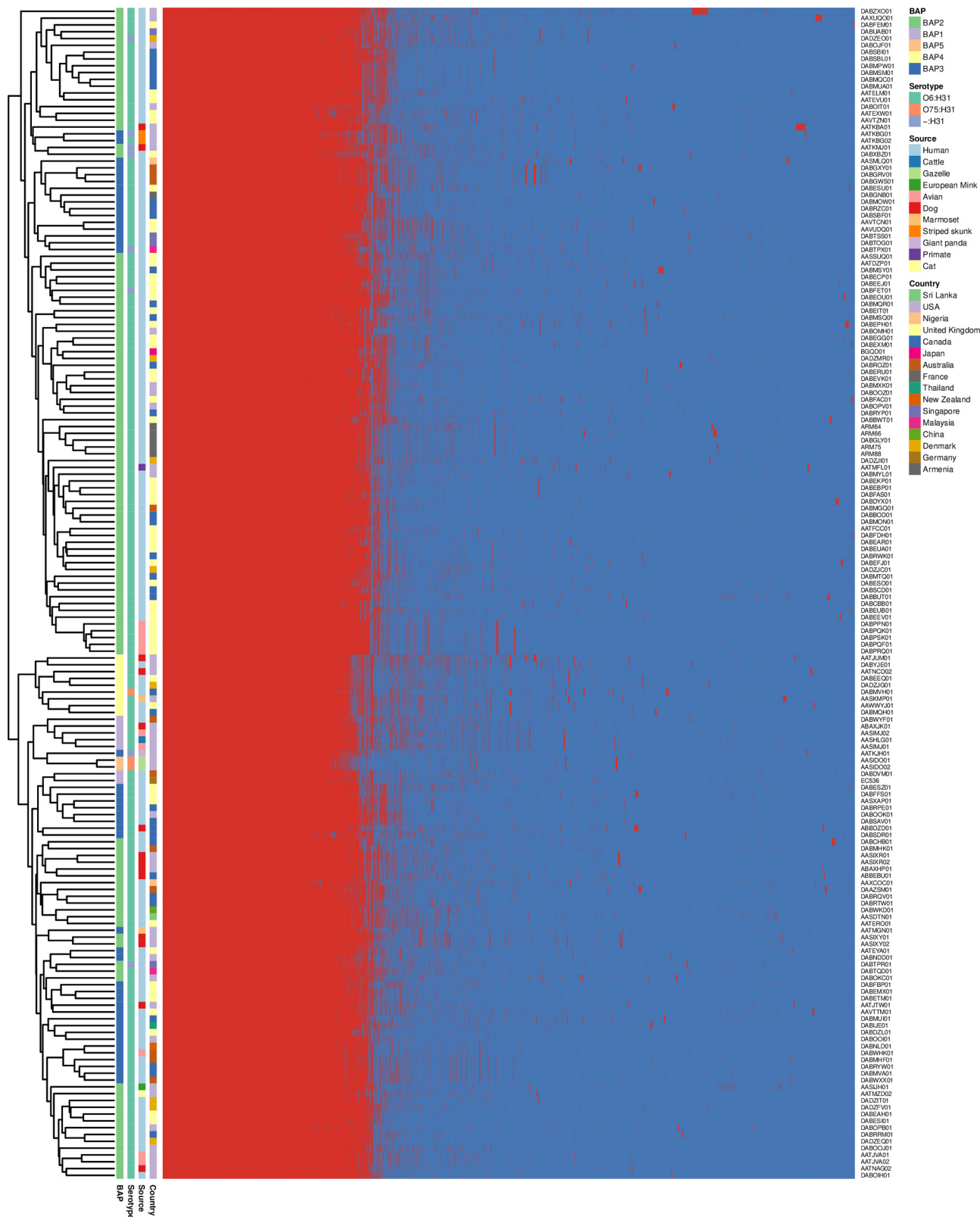


FIG 3 Hierarchy cluster heat map of the accessory genome of ST127 *E. coli* isolates.

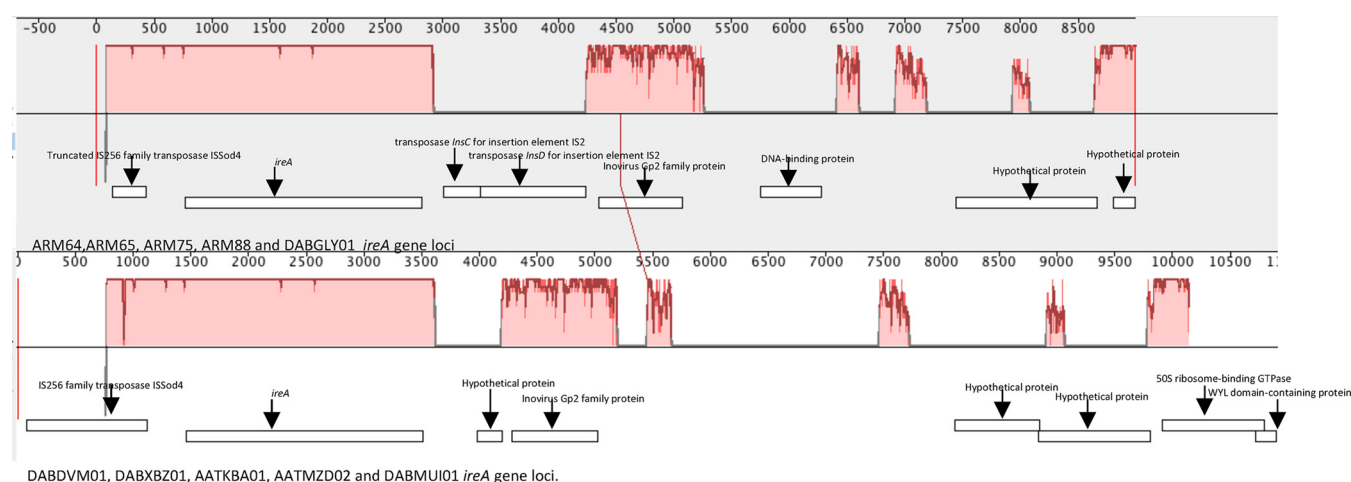


FIG 5 Mauve output of genome loci comparison of the *ireA* gene. The red block indicates genome similarities between the two loci.

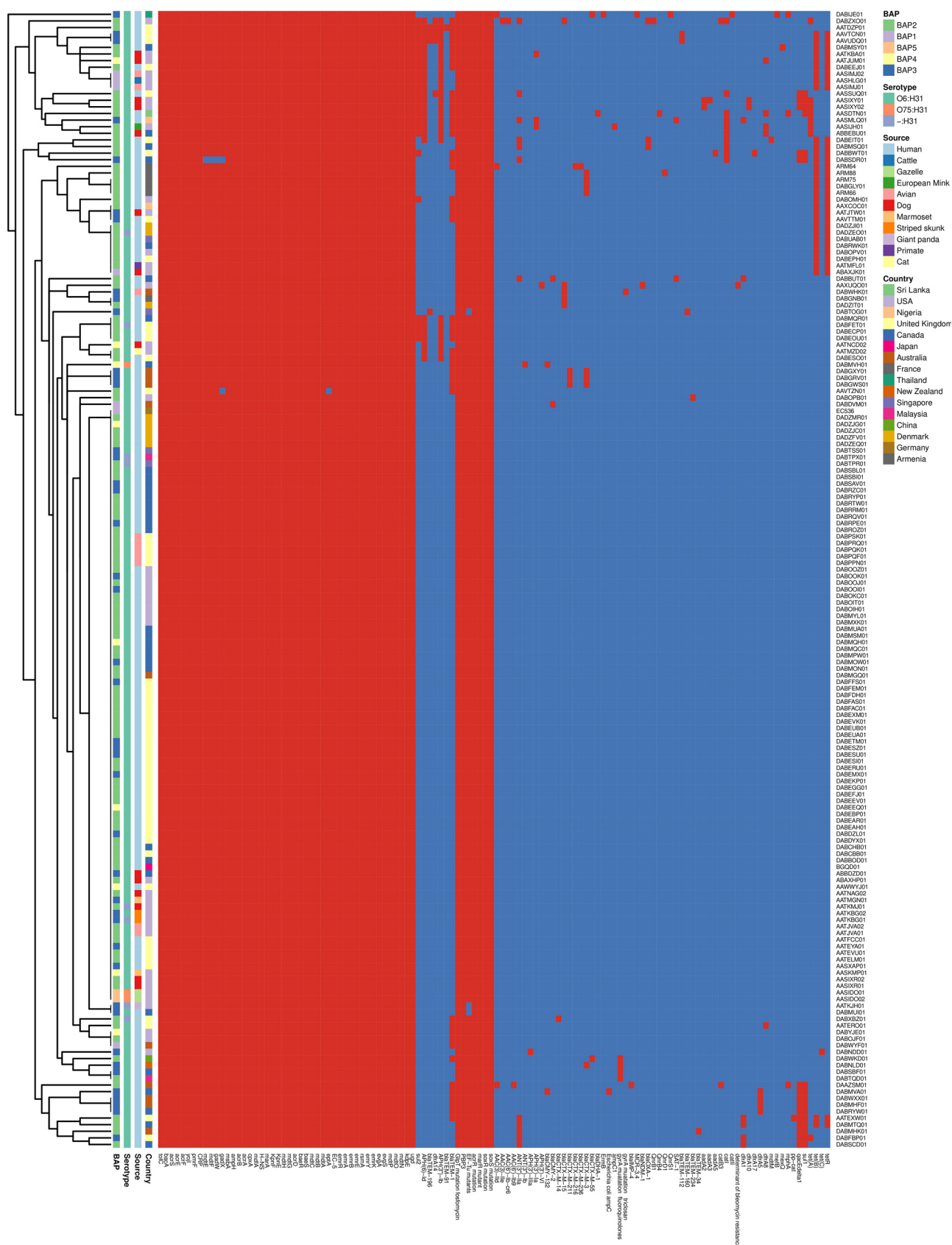
the isolates recovered in our study and DABGLY01 harbored two additional insertion sequence transposases instead of a hypothetical protein gene. In addition, the Inovirus Gp2 protein family gene found in the *ireA* locus shared 87.24% protein similarity (Fig. 5).

To determine whether the unique genes and 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 feces 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, the United States of America, France, Switzerland, and the United Kingdom and from mammalian, avian, human, and wastewater sources. However, we did not find isolates with $\geq 99.99\%$ homology that contained all three genomic regions; instead, we detected multiple isolates that harbored the *ireA* and *hlyA* genomic region and the *ireA* and *lutABC/hsdMSR* genomic region. However, all three genomic regions were identified in various isolates belonging to the ST117 sequence type.

Antibiotic resistance genotype. The antimicrobial resistance genes and mutations of all of the Armenian *E. coli* ST127 isolates were compared to 168 WGS data sets 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 (Fig. 6). 47 of these genes/mutations were present in all 172 of the *E. coli* ST127 isolates that we analyzed. These core antibiotic resistance genes/mutations are known to induce resistance to antibiotics, including those belonging to the aminocoumarin, aminoglycoside, carbapenem, cephalosporin, cephamycin, diaminopyrimidine, fluoroquinolone, fosfomycin, glycylicycline, macrolide, monobactam, nitroimidazole, nucleoside, penam, peptide, phenicol, rifamycin, and tetracycline classes.

All four of the isolates that were recovered in this study possessed various degrees of antibiotic resistance genes/mutations, including 1 isolate (ARM64) possessing 59 antibiotic resistance genes/mutations, 1 isolate (ARM88) possessing 57 antibiotic resistance genes/mutations, and 2 isolates (ARM66 and ARM75) possessing 56 antibiotic resistance genes mutations. These numbers are greater than the average number of antibiotic resistance genes, compared to the *E. coli* ST127 isolates obtained from the ENA archive (average: 55 genes; range: 52 to 67 genes) that have been reported previously. In addition, a plasmid-mediated quinolone resistance protein gene *qnrD1* was unique to the ARM88 isolate recovered in this study.

60 out of the 172 (34.88%) ST127 isolates, and all 4 of the Armenian isolates that were recovered in this study, possessed genes that encode ESBL, including but not limited to:



CTX ($n = 17$), TEM ($n = 48$), and OXA-1 ($n = 3$). The most common CTX gene was *bla*_{CTX-M-3} ($n = 16$), which was also found in 3 of the Armenian isolates (ARM66, ARM75, and ARM88), *bla*_{CTX-M-15} ($n = 15$), *bla*_{CTX-M-211} ($n = 6$), *bla*_{CTX-M-55} ($n = 4$), *bla*_{CTX-M-216} and *bla*_{CTX-M-236} ($n = 2$), which was found in one of the Armenian isolates (ARM64), and *bla*_{CTX-M-14} ($n = 2$). The most common TEM gene was *bla*_{TEM-1} ($n = 45$), which was also found in ARM64, *bla*_{TEM-196} ($n = 4$), *bla*_{TEM-112} ($n = 2$), *bla*_{TEM-160} ($n = 1$), *bla*_{TEM-234} ($n = 1$), *bla*_{TEM-34} ($n = 1$), and *bla*_{TEM-91} ($n = 1$). In addition, the carbapenem-resistance genes *bla*_{NMD-1} ($n = 1$) and *bla*_{IPM-4} ($n = 1$) were detected in one of each of the *E. coli* ST127 isolates. However, they were detected only in those obtained from the ENA archive. We also found the antibiotic resistance genes/mutations, which were found in nearly all of the *E. coli* ST127 isolates from the ENA archive and in the all of the Armenian *E. coli* ST127 isolates, including the multidrug efflux *mdtE*, *mdtF*, the *gadW* and *gadX*, the pmr phosphoethanolamine transferase gene *eptA* (found in 171 of the 172 isolates, except for *gadX*, which was found in 170 of the 172 isolates), the elfamycin antibiotic resistance EF-Tu mutation (169 of the 172 *E. coli* ST127 isolates) and the tetracycline efflux pump genes *tet(B)* and *tetR* (32 of the 172 *E. coli* ST127 isolates). In addition, the aminoglycoside acetyltransferase gene *AAC(3)-Ild* was identified in ARM64 (3 of the 172 *E. coli* ST127 isolates).

Plasmid typing. A plasmid replicon was identified in 72% (124/172) of the *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 (26) 1_pUT189 ($n = 48$) (Fig. 7). The greatest 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 ($n = 31$). The ARM64 isolates harbored 3 plasmid replicons, including IncFII_1 (found in 8.14% of the ST127 isolates), Inc1_1_Alpha (found in 5.23% of the ST127 isolates), and IncX1 (found in 4.07% of the ST127 isolates). Moreover, ARM66 harbored one plasmid replicon, Inc12_1, that was unique to it. Overall, we found that isolates, based on the hierarchy cluster heat map 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 the DNA sequencing of pathogens that cause 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 a whole-genome analysis of ESBL UPEC ST127 isolates recovered from hospitalized patients in Armenia, 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 the Armenian isolates. Interestingly, all of the Armenian *E. coli* ST127 isolates were identified as UPEC ESBL and belonged to the same genetic lineage. Despite these isolates being recovered from patients in five different hospitals, their inferred divergence date of the most recent common ancestor was around 2000 (95% CI of the date: 1995 to 2006). The close genetic similarities of the Armenian isolates indicated that ST127 was potentially a dominant lineage in causing urinary tract infections in Armenia at the time of this study (January to September of 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 that were near known prophage genes in both the Armenian and French isolate clusters, which suggested 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 that were recovered from different human (including human urine samples) and animal sources. However, none of these isolates were shown to harbor all three genomic regions, suggesting a unique combination of genes that is 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.

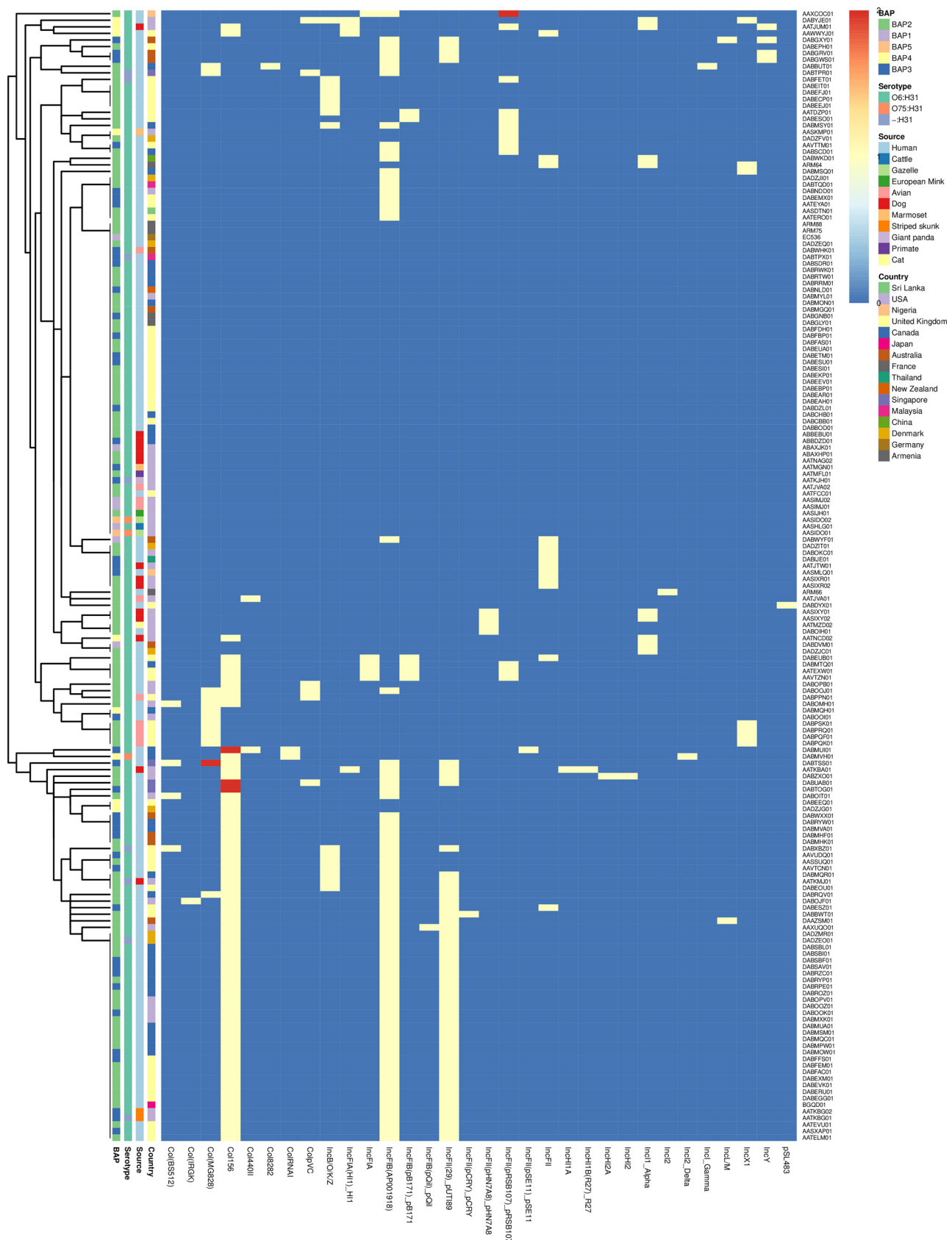


FIG 7 Hierarchy cluster heat map of ST127 *E. coli* plasmid replicon.

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 hemolytic action toward mammalian cells in anaerobic conditions (27). 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, 28–31). Moreover, we found that the haemolysin A gene *hlyA*, which is unrelated to *hlyE*, was ubiquitous in the Armenian *E. coli* ST127 isolates and has previously been described as an important virulence factor in causing pyelonephritis (27, 32, 33). Therefore, it is plausible that both genes can act synergistically to cause hemolysis in urinary tract infections; however, no such data have been reported up to date due to the rarity of isolates harboring both the *hlyA* and *hlyE* genes (29, 30). The *hlyE* gene region was determined to be a part of a prophage insertion region using the Prophage Hunter software package (34). We found that a small portion of the prophage region was relatively similar to an Stx2-converting phage Stx2a_F349, which is well-known for its horizontal transfer of the Shiga toxin (Stx) in Stx-producing *E. coli* (STEC). However, scientific evidence obtained under 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 (35). Nevertheless, the BLAST search analysis of these genes in the *hlyE* prophage region identified two genes that were associated with prophages but did not originate from an Stx2-converting phage. These genes have shown homology to the Inovirus-type Gp2 family protein and the transcriptional regulator gene *alpA*, which is associated with the cryptic P4-like prophage CP4-57. Therefore, we can hypothesize that the prophage insertion region of *hlyE* originated from an unknown prophage species.

The 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 (considered to be unique if they exhibited less than <90% protein homology) and other unique genes as being adjacent to prophage genes in the bacterial chromosome. For example, we identified another gene that had homology to an Inovirus Gp2 family protein that was unique to our isolates as well as the IS3 family transposase IS2 near the TonB-dependent siderophore receptor gene *ireA*, which is involved in the increased uptake of iron within an iron-restricted environment, such as human urine (36). The *ireA* was found in five other isolates that were retrieved from ENA and were recovered from companion animals and human urine samples, but these isolates 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 the protein level. The difference in homology may be linked to the high mutation rates within filamentous phages due to them having single-stranded DNA genomes (37). We also found a difference between the genome regions of the Armenian and French isolates cluster, as they had insertion sequence IS3 family transposase IS2 elements between the *ireA* and Inovirus Gp2 family protein instead of a hypothetical protein, as found within the other ST127 *ireA* isolates. Moreover, a truncated version of the 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 immobilization and stability of the prophage region, which may suggest why we did not find other *E. coli* ST127 clusters containing the *ireA* gene (38, 39).

Furthermore, the Armenian and French isolates cluster had a genomic region that contained 11 unique genes that were adjacent to the prophage integrase gene *intA* and the tRNA gene. Downstream of these genes were the unique genes to the Armenian and French cluster, including a peptide ABC transporter substrate-binding protein, the lactate utilization 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 the involvement of these genes in UPEC growth and survival in the urinary tract. Nonetheless, we can hypothesize that the peptide ABC transporter substrate-binding protein gene, *lutABC* operon, and *lutP* permease, may have a role in utilizing lactate as an energy source, along with other metabolites

found in the urine; however, further studies are required (40–42). In addition, the *hsdMSR* operon may have initially acted as a superinfection exclusion mechanism to prevent secondary bacteriophage infections (43). Moreover, the *hsdMSR* operon may act as a gene regulation mechanism within the urinary tract via DNA methylation (44, 45).

Although we were not able to find a complete arsenal of prophage gene integration within all of these genomic regions, we hypothesize that the most likely mode of horizontal transfer for 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 remnants of prophage genes (46). Furthermore, based on the date of divergence (1966 date CI: 1956 to 1977) of the United Kingdom isolates recovered from blood cultures in 2011, which were the closest related isolates to the Armenian and French cluster, we cannot determine when these genomic regions integrated into the Armenian and French *E. coli* ST127 lineage or whether it happened before or after they diverged from the United Kingdom isolates.

Although we did not find antibiotic resistance gene mutations and plasmid replicons that were ubiquitous and unique to the Armenian isolates, we did identify that the quinolone resistance protein gene *qnrD1* was unique to ARM88 and that 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 carbapenem and colistin antibiotic resistance genes (47–52). However, we could not link them to any plasmids because we could not identify a known plasmid replicon with an ARM88 plasmid sequence. In addition, we did not detect any antibiotic resistance gene that was associated with carbapenem and colistin antibiotic resistance in either ARM64 or ARM66. In correlation with the phenotypic data, we determined that ARM88 *qnrD1* showed known resistance to fluoroquinolone antibiotics (norfloxacin and levofloxacin). Intriguingly, we found 46 antibiotic resistance determinants that were ubiquitous in all of the *E. coli* ST127 isolates that were studied, which coincides with an analysis by Goldstone and Smith (2017) that described 50 antibiotic resistance genes, which may play a role in providing basal-level resistance toward a diverse number of antimicrobial compounds, as being part of the core resistome ($\geq 95\%$ of the isolates) (53). However, Goldstone and Smith (53) 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 they have been involved in antibiotic resistance in other bacterial species. Our findings indicate that the β -lactam and cephalosporin antibiotic resistance phenotype in the Armenian isolates can be attributed to the ESBL associated genes *bla*_{CTX-M-3}, *bla*_{CTX-M-236}, and *bla*_{TEM-1}. However, we were unable to determine the antibiotic resistance genotype that confers phenotypic resistance to the aminoglycoside antibiotic amikacin in ARM88 resistance to the fluoroquinolone antibiotics norfloxacin and levofloxacin in ARM75, or intermediate resistance to the carbapenem antibiotics as well as imipenem resistance in ARM66. Moreover, the antibiotic resistance phenotypes observed in our study were consistent with the findings of previous studies that report that ST127 are sensitive toward a wide range of antibiotics, compared to other UPEC sequence types, and that the aminoglycoside and fluoroquinolone antibiotic resistances as well as ESBL production were not commonly found in ST127 isolates (5, 54, 55).

The main limitation of this study is its 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 share a recent common ancestor. In addition, all of the isolates shared virulence and metabolic genes that were acquired via horizontal transfer and 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 and the roles of some of these genes as well as 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 recognize borders.

MATERIALS AND METHODS

Bacterial isolation and identification. 12 *E. coli* isolates were received from the medical microbiology laboratories of 5 hospitals in Armenia between January and August of 2019. All isolates were recovered from urine specimens of hospitalized patients. The isolates were identified as *E. coli* using matrix-assisted laser desorption/ionization-time of flight mass spectrometry (MALDI-TOF-MS), as described previously (56). 4 out of 12 isolates that were identified as *E. coli* ST127 were selected for the purposes of this study.

Antibiotic susceptibility testing. All 12 isolates were tested for their antibiotic susceptibility to a panel of 11 antibiotics, including ampicillin (10 mg), piperacillin-tazobactam (30/6 mg), amoxicillin and clavulanic acid (20/10 mg), ceftazidime (10 mg), cefepime (30 mg), norfloxacin (10 mg), levofloxacin (5 mg), amikacin (30 mg), imipenem (10 mg), meropenem (10 mg), and chloramphenicol (30 mg) (Mast Group, Merseyside, UK), using a disk diffusion method, according to the European Committee on Antimicrobial Susceptibility Testing protocol (57). The antibiotics chosen were those that are the most frequently used in clinical settings in Armenia. *E. coli* isolates were identified as “ESBL-producing” upon the confirmation of their resistance to cefepime and ceftazidime antibiotics.

Genome sequencing and assembly. All of the *E. coli* isolates were whole-genome sequenced using an Illumina HiSeq platform. However, for the purposes of this study, WGS analyses were conducted for only 4 isolates belonging to ST127. Genomic DNA was extracted using a TIANamp Bacteria DNA Kit (Tiangen, China) and the paired-end sequencing libraries were constructed using Nextera XT DNA Sample Preparation Kits or a TruSeq DNA HT Sample Prep Kit (Illumina, USA), following the manufacturers’ instructions.

The quality of the short-reads was analyzed using fastQC, and low quality reads were trimmed using the Trimmomatic software package (58). The trimmed reads were *de novo* assembled using SPAdes (26).

Genome selection for phylogenetic and genomic comparison. To conduct a comparative genomic analysis of the *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 the date and source of isolation and the country of origin. To select only the ST127 genomes among the 12 *E. coli* isolates recovered in this study and those retrieved from the ENA database, we screened the draft *E. coli* genomes using mlst (<https://github.com/tseemann/mlst>, accessed July 2021) with the Achtman typing scheme from pubMLST (accessed January 2022).

Phylogenetic analysis. To construct a core SNP maximum-likelihood (ML) phylogenetic tree of the *E. coli* ST127 isolates, we first aligned all of the isolates against the reference *E. coli* ST127 genome EC536 (accession no. [NC_008253.1](https://ncbi.nlm.nih.gov/nuccore/NC_008253.1)) using the Parsnp alignment software from the Harvest suite (59). Then, recombination was removed from the aligned sequences using Gubbins, and a phylogenetic tree was constructed from the recombination-free alignment using IQtree v2.1.2 (60). To select the best model for the construction of the phylogenetic tree, we used ModelFinder and set the ultrafast bootstrap replication to 1,000 (61, 62). Finally, the phylogenetic tree was visualized using iTOL (63). The SNP-distances between the isolates core genomes were worked out from the recombination-free alignment using the SNP-dist software package (<https://github.com/tseemann/snp-dists>).

A Bayesian dated maximum clade credibility (MCC) tree was reconstructed using BEAST v1.10.4 (64). A GTR empirical substitution model, as determined by ModelFinder, was used for the Bayesian analysis and was set with a Bayesian skyline strict clock model with the MCC chain set to 100 million. 2 independent runs trees were combined using LogCombiner with a 10% burn-in and an MCC tree was constructed from the combined trees using TreeAnnotator. The MCC tree was visualized using the FigTree software package (<http://tree.bio.ed.ac.uk/software/figtree/>).

Genome annotation. The genomes of the *E. coli* ST127 isolates were annotated using Prokka (65). Virulent gene markers and plasmid replicons were screened using the Abricate software package (<https://github.com/tseemann/abricate>) in conjunction with the VirulenceFinder and Virulence Factor Database combined for virulence detection and the PlasmidFinder database for the identification of plasmid replicons (accessed January 2022) (66–68). Antibiotic resistance genes/mutations were screened using the Resistance Gene Identifier software in conjunction with the Comprehensive Resistant Antibiotic Resistance Database (accessed January 2022) (69). The isolates were serotyped by ECTyper (70).

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 to 90% (71). Unique genes within the different *E. coli* ST127 groups were identified using Scoary (72). Unique genes and genes within their loci were further annotated via a BLAST against the UniProt database and the NCBI nucleotide collection. MAFFT was used to identify individual protein percentage similarities (73). The genomic region comparison was visualized using Mauve v 2.4.0 (74).

The MOB-suite was used to reconstruct plasmid sequences from draft genomes to determine whether unique genes were plasmid-borne (75). Integrative and conjugative element genomic regions were detected using ICEfinder with the ICEberg 2.0 database (76). Prophage Hunter was used to identify the prophage regions in the draft genomes to determine whether unique genes were within the prophage insertion region (34). The horizontal gene transfer regions that contained these unique genes were compared to other *E. coli* isolates via a sequence similarity search using NCBI blast against the ENA sequence archive (77).

R programs. The grouping of isolates into subclusters within the SNP core whole-genome alignment was conducted using the hierarchical clustering R package “RhierBAPS” (78). In addition, a hierarchy cluster heat map of antibiotic resistant genes/mutations, accessory genes, and plasmid replicons was constructed using the pheatmap R package (<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](https://www.ebi.ac.uk/ena/record/PRJEB51925). Individual sequence file accession numbers are included in Table S2.

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

SUPPLEMENTAL FILE 1, XLSX file, 0.04 MB.

ACKNOWLEDGMENTS

We declare no potential conflicts of interest.

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

REFERENCES

- 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. <https://doi.org/10.1038/nrmicro3432>.
- Kot B. 2019. Antibiotic resistance among uropathogenic *Escherichia coli*. *Pol J Microbiol* 68:403–415. <https://doi.org/10.33073/pjm-2019-048>.
- Church LWP. 2008. Enterobacteriaceae, p 945–952. In Schlossberg D (ed), *Clinical Infectious Disease*. Cambridge University Press, Cambridge.
- Sarowska J, Futoma-Koloch B, Jama-Kmiecik A, Frej-Madrzak M, Ksiaczczyk 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. <https://doi.org/10.1186/s13099-019-0290-0>.
- 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. *J Antimicrob Chemother* 67:346–356. <https://doi.org/10.1093/jac/ckr451>.
- 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. <https://doi.org/10.1128/CMR.00135-18>.
- 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. <https://doi.org/10.1371/journal.pone.0244358>.
- 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. <https://doi.org/10.3389/fmicb.2020.02065>.
- 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. <https://doi.org/10.1371/journal.pone.0179880>.
- 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. *Front Microbiol* 7:173. <https://doi.org/10.3389/fmicb.2016.00173>.
- 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. <https://doi.org/10.1111/j.1574-6976.2008.00136.x>.
- Rossolini GM, D'Andrea MM, Mugnaioli C. 2008. The spread of CTX-M-type extended-spectrum β -lactamases. *Clin Microbiol Infect* 14:33–41. <https://doi.org/10.1111/j.1469-0691.2007.01867.x>.
- Bezabih YM, Bezabih A, Dion M, Batard E, Tekla 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 Antimicrob Resist* 4:dlac048. <https://doi.org/10.1093/jacamr/dlac048>.
- Devi LS, Broor S, Rautela RS, Grover SS, Chakravarti A, Chattopadhyaya D. 2020. Increasing prevalence of *Escherichia coli* and *Klebsiella pneumoniae* producing CTX-M-type extended-spectrum β -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. https://doi.org/10.4103/ijabmr.IJABMR_360_19.
- Martischang R, François P, Cherkaoui A, Gaia 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. *Antimicrob Resist Infect Control* 10:148. <https://doi.org/10.1186/s13756-021-01013-7>.
- Elankumaran P, Browning GF, Marenda MS, Reid CJ, Djordjevic SP. 2022. Close genetic linkage between human and companion animal extraintestinal pathogenic *Escherichia coli* ST127. *Curr Res Microb Sci* 3:100106. <https://doi.org/10.1016/j.crmicr.2022.100106>.
- 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. <https://doi.org/10.1186/s12941-021-00470-7>.
- 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. *Front Microbiol* 9:1217. <https://doi.org/10.3389/fmicb.2018.01217>.
- Johnson JR, Russo TA. 2018. Molecular epidemiology of extraintestinal pathogenic *Escherichia coli*. *EcoSal Plus* 8. <https://doi.org/10.1128/ecosalplus.ESP-0004-2017>.
- 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 β -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>.
- 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. *Antimicrob Resist Infect Control* 10:63. <https://doi.org/10.1186/s13756-021-00931-w>.
- Shen H, Ma H. 2013. Molecular epidemiology in low- and middle-income countries, p 157–173. In Soliman A, Schottenfeld D, Boffetta P (ed), *Cancer Epidemiology*. Oxford University Press.
- Data for Armenia, Upper middle income. Data. World Bank. <https://data.worldbank.org/locations=AM-XT>. Retrieved 18 May 2022.
- 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. <https://doi.org/10.1186/s13756-017-0169-0>.
- Jamroz 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. *Front Microbiol* 10. <https://doi.org/10.3389/fmicb.2019.01592>.
- Prijbelski A, Antipov D, Meleshko D, Lapidus A, Korobeynikov A. 2020. Using SPAdes de novo assembler. *Curr Protoc Bioinformatics* 70:e102. <https://doi.org/10.1002/cpbi.102>.
- Hunt S, Green J, Artymuk PJ. 2010. Hemolysin E (HlyE, ClyA, SheA) and related toxins, p 116–126. In Anderlüh G, Lakey J (ed), *Proteins Membrane Binding and Pore Formation*. Springer, New York, NY.
- Hutton TA, Innes GK, Harel J, Garneau P, Cucchiara A, Schifferli DM, Rankin SC. 2018. Phylogroup and virulence gene association with clinical characteristics of *Escherichia coli* urinary tract infections from dogs and cats. *J Vet Diagn Invest* 30:64–70. <https://doi.org/10.1177/1040638717729395>.
- Kerényi M, Allison HE, Batai I, Sonnevend Á, Emödy L, Plaveczyk N, Pál T. 2005. Occurrence of hlyA and sheA genes in extraintestinal *Escherichia coli* strains. *J Clin Microbiol* 43:2965–2968. <https://doi.org/10.1128/JCM.43.6.2965-2968.2005>.
- Lorenz SC, Son I, Maounounen-Laasri A, Lin A, Fischer M, Kase JA. 2013. Prevalence of hemolysin genes and comparison of ehxA subtype patterns in Shiga toxin-producing *Escherichia coli* (STEC) and non-STEC strains from clinical, food, and animal sources. *Appl Environ Microbiol* 79: 6301–6311. <https://doi.org/10.1128/AEM.02200-13>.
- Massella E, Giacometti F, Bonilauri P, Reid CJ, Djordjevic SP, Meriardi G, Bacci C, Fiorentini L, Massi P, Bardasi L, Rubini S, Savini F, Serraino A, Piva S. 2021. Antimicrobial resistance profile and ExPEC virulence potential in commensal *Escherichia coli* of multiple sources. *Antibiotics* 10:351. <https://doi.org/10.3390/antibiotics10040351>.

32. Ristow LC, Welch RA. 2016. Hemolysin of uropathogenic *Escherichia coli*: a cloak or a dagger? *Biochim Biophys Acta* 1858:538–545. <https://doi.org/10.1016/j.bbame.2015.08.015>.
33. Wang C, Li Q, Lv J, Sun X, Cao Y, Yu K, Miao C, Zhang Z-S, Yao Z, Wang Q. 2020. Alpha-hemolysin of uropathogenic *Escherichia coli* induces GM-CSF-mediated acute kidney injury. *Mucosal Immunol* 13:22–33. <https://doi.org/10.1038/s41385-019-0225-6>.
34. Song W, Sun H-X, Zhang C, Cheng L, Peng Y, Deng Z, Wang D, Wang Y, Hu M, Liu W, Yang H, Shen Y, Li J, You L, Xiao M. 2019. Prophage Hunter: an integrative hunting tool for active prophages. *Nucleic Acids Res* 47: W74–W80. <https://doi.org/10.1093/nar/gkz380>.
35. Tozzoli R, Grande L, Michelacci V, Ranieri P, Maugliani A, Caprioli A, Morabito S. 2014. Shiga toxin-converting phages and the emergence of new pathogenic *Escherichia coli*: a world in motion. *Front Cell Infect Microbiol* 4:80. <https://doi.org/10.3389/fcimb.2014.00080>.
36. Torres-Puig S, García V, Stærk K, Andersen TE, Møller-Jensen J, Olsen JE, Herrero-Fresno A. 2022. “Omics” technologies - what have they told us about uropathogenic *Escherichia coli* fitness and virulence during urinary tract infection? *Front Cell Infect Microbiol* 12:824039. <https://doi.org/10.3389/fcimb.2022.824039>.
37. Peck KM, Luring AS. 2018. Complexities of viral mutation rates. *J Virol* 92. <https://doi.org/10.1128/JVI.101031-17>.
38. Ooka T, Ogura Y, Asadulghani M, Ohnishi M, Nakayama K, Terajima J, Watanabe H, Hayashi T. 2009. Inference of the impact of insertion sequence (IS) elements on bacterial genome diversification through analysis of small-size structural polymorphisms in *Escherichia coli* O157 genomes. *Genome Res* 19:1809–1816. <https://doi.org/10.1101/gr.089615.108>.
39. Ramisetty BCM, Sudhakari PA. 2019. Bacterial ‘grounded’ prophages: hot-spots for genetic renovation and innovation. *Front Genet* 10:65. <https://doi.org/10.3389/fgene.2019.00065>.
40. Hua Q, Joyce AR, Palsen BØ, Fong SS. 2007. Metabolic characterization of *Escherichia coli* strains adapted to growth on lactate. *Appl Environ Microbiol* 73:4639–4647. <https://doi.org/10.1128/AEM.00527-07>.
41. Mann R, Mediati DG, Duggin IG, Harry EJ, Bottomley AL. 2017. Metabolic adaptations of uropathogenic *E. coli* in the urinary tract. *Front Cell Infect Microbiol* 7:241. <https://doi.org/10.3389/fcimb.2017.00241>.
42. 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. <https://doi.org/10.1006/bbrc.2001.6255>.
43. 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, Isem S, Johnson AA, Kenna MA, Klyczek KK, Magee 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. <https://doi.org/10.1038/nmicrbiol.2016.251>.
44. Casadesús J, Low D. 2006. Epigenetic gene regulation in the bacterial world. *Microbiol Mol Biol Rev* 70:830–856. <https://doi.org/10.1128/MMBR.00016-06>.
45. Murray NE. 2000. Type I restriction systems: sophisticated molecular machines (a legacy of Bertani and Weigle). *Microbiol Mol Biol Rev* 64: 412–434. <https://doi.org/10.1128/MMBR.64.2.412-434.2000>.
46. Bobay L-M, Touchon M, Rocha EPC. 2014. Pervasive domestication of defective prophages by bacteria. *Proc Natl Acad Sci U S A* 111:12127–12132. <https://doi.org/10.1073/pnas.1405336111>.
47. Albornoz E, Lucero C, Romero C, Quiroga MP, Rapoport M, Guerriero L, Andres P, Rodriguez C, Galas M, Centron D, Corso A, Petroni A, WHONET-Argentina Group. 2017. Prevalence of plasmid-mediated quinolone resistance genes in clinical Enterobacteria from Argentina. *Microb Drug Resist* 23:177–187. <https://doi.org/10.1089/mdr.2016.0033>.
48. Liu Z, Xiao X, Li Y, Liu Y, Li R, Wang Z. 2019. Emergence of IncX3 plasmid-harboring bla_{NDM-5} dominated by *Escherichia coli* ST48 in a goose farm in Jiangsu, China. *Front Microbiol* 10. <https://doi.org/10.3389/fmicb.2019.02002>.
49. 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. <https://doi.org/10.1016/j.jmii.2016.07.004>.
50. Meinersmann RJ. 2019. The biology of IncI2 plasmids shown by whole-plasmid multi-locus sequence typing. *Plasmid* 106:102444. <https://doi.org/10.1016/j.plasmid.2019.102444>.
51. 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. <https://doi.org/10.2147/IDR.S210554>.
52. 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. <https://doi.org/10.1128/mSphere.00917-20>.
53. Goldstone RJ, Smith DGE. 2017. A population genomics approach to exploiting the accessory “resistome” of *Escherichia coli*. *Microb Genom* 3: e000108. <https://doi.org/10.1099/mgen.0.000108>.
54. 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. <https://doi.org/10.1186/1757-4749-6-32>.
55. Fibke CD, Croxen MA, Geum HM, Glass M, Wong E, Avery BP, Daignault D, Mulvey MR, Reid-Smith RJ, Parnley 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. <https://doi.org/10.1093/ofid/ofz431>.
56. Mkrtchyan HV, Russell CA, Wang N, Cutler RR. 2013. Could public restrooms be an environment for bacterial resistomes? *PLoS One* 8:e54223. <https://doi.org/10.1371/journal.pone.0054223>.
57. The European Committee on Antimicrobial Susceptibility Testing. Breakpoint tables for interpretation of MICs and zone diameters. Version 11.0, 2021. <http://www.eucast.org>.
58. Babraham Bioinformatics. FastQC: a quality control tool for high throughput sequence data. <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>. Retrieved 16 November 2021.
59. 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 Biol* 15:524. <https://doi.org/10.1186/s13059-014-0524-x>.
60. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* 32:268–274. <https://doi.org/10.1093/molbev/msu300>.
61. Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. 2018. UFBoot2: improving the ultrafast bootstrap approximation. *Mol Biol Evol* 35:518–522. <https://doi.org/10.1093/molbev/msx281>.
62. Kalyanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermini LS. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods* 14:587–589. <https://doi.org/10.1038/nmeth.4285>.
63. Letunic I, Bork P. 2021. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res* 49:W293–W296. <https://doi.org/10.1093/nar/gkab301>.
64. Suchard MA, Lemey P, Baele G, Ayres DL, Drummond AJ, Rambaut A. 2018. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evol* 4:vey016. <https://doi.org/10.1093/ve/vey016>.
65. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30:2068–2069. <https://doi.org/10.1093/bioinformatics/btu153>.
66. 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* 58:3895–3903. <https://doi.org/10.1128/AAC.02412-14>.
67. Chen L, Yang J, Yu J, Yao Z, Sun L, Shen Y, Jin Q. 2005. VFDB: a reference database for bacterial virulence factors. *Nucleic Acids Res* 33:D325–328. <https://doi.org/10.1093/nar/gki008>.
68. Joensen KG, Scheutz F, Lund O, Hasman H, Kaas RS, Nielsen EM, Aarestrup FM. 2014. Real-time whole-genome sequencing for routine typing, surveillance, and outbreak detection of verotoxigenic *Escherichia coli*. *J Clin Microbiol* 52:1501–1510. <https://doi.org/10.1128/JCM.03617-13>.
69. Alcock BP, Raphenya AR, Lau TTY, Tsang KK, Bouchard M, Edalatmand A, Huynh V, Nguyen A-LV, Cheng AA, Liu S, Min SY, Miroshnichenko A, Tran H-K, Werfalli RE, Nasir JA, Oloni M, Speicher DJ, Florescu A, Singh B, Faltyn M, Hernandez-Koutoucheva A, Sharma AN, Bordeleau E, Pawlowski AC, Zubyk HL, Dooley D, Griffiths E, Maguire F, Winsor GL, Beiko RG, Brinkman FSL, Hsiao WWL, Domselaar GV, McArthur AG. 2020. CARD 2020: antibiotic resistance surveillance with the comprehensive antibiotic resistance database. *Nucleic Acids Res* 48:D517–D525.
70. Bessonov K, Laing C, Robertson J, Yong I, Ziebell K, Gannon VPJ, Nichani A, Arya G, Nash JHE, Christianson S. 2021. ECTyper: in silico *Escherichia coli* serotype and species prediction from raw and assembled whole-genome sequence data. *Microb Genom* 7.
71. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MTG, Fookes M, Falush D, Keane JA, Parkhill J. 2015. Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics* 31:3691–3693. <https://doi.org/10.1093/bioinformatics/btv421>.

72. Brynildsrud O, Bohlin J, Scheffer L, Eldholm V. 2016. Rapid scoring of genes in microbial pan-genome-wide association studies with Scoary. *Genome Biol* 17:238. <https://doi.org/10.1186/s13059-016-1108-8>.
73. Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 30:772–780. <https://doi.org/10.1093/molbev/mst010>.
74. Darling ACE, Mau B, Blattner FR, Perna NT. 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res* 14:1394–1403. <https://doi.org/10.1101/gr.2289704>.
75. Robertson J, Nash JHE. 2018. MOB-suite: software tools for clustering, reconstruction and typing of plasmids from draft assemblies. *Microb Genom* 4.
76. Liu M, Li X, Xie Y, Bi D, Sun J, Li J, Tai C, Deng Z, Ou H-Y. 2019. ICEberg 2.0: an updated database of bacterial integrative and conjugative elements. *Nucleic Acids Res* 47:D660–D665. <https://doi.org/10.1093/nar/gky1123>.
77. Madeira F, Park YM, Lee J, Buso N, Gur T, Madhusoodanan N, Basutkar P, Tivey ARN, Potter SC, Finn RD, Lopez R. 2019. The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res* 47:W636–W641. <https://doi.org/10.1093/nar/gkz268>.
78. Tonkin-Hill G, Lees JA, Bentley SD, Frost SDW, Corander J. 2018. RhierBAPS: an R implementation of the population clustering algorithm hierBAPS. *Wellcome Open Res* 3:93. <https://doi.org/10.12688/wellcomeopenres.14694.1>.