



Arrangements of Mobile Genetic Elements among Virotype E Subpopulation of *Escherichia coli* Sequence Type 131 Strains with High Antimicrobial Resistance and Virulence Gene Content

Omid Pajand,^{a,b} Hamzeh Rahimi,^c Narges Darabi,^d Solaleh Roudi,^e Khatereh Ghassemi,^d Frank M. Aarestrup,^a Pimlapas Leekitcharoenphon^a

^aSection for Genomic Epidemiology, National Food Institute, Technical University of Denmark, Lyngby, Denmark

^bMicrobiology Department, Faculty of Medicine, Semnan University of Medical Sciences, Semnan, Iran

^cDepartment of Molecular Medicine, Biotechnology Research Center, Pasteur Institute of Iran, Tehran, Iran

^dStudent Research Committee, Faculty of Medicine, Semnan University of Medical Sciences, Semnan, Iran

^eImam Hossein Hospital, Semnan University of Medical Sciences, Aradan, Iran

ABSTRACT *Escherichia coli* sequence type 131 (ST131) is known for its contribution to multidrug resistance and the worldwide spread of this clone has become a global problem. Understanding the trends among ST131 clades will help design strategies to prevent its rapid dissemination. In this study, 72 ST131 strains were subjected to comparative genomic analysis and 64 clade C strains were compared with clade C strains reported from other regions using publicly available whole-genome sequencing data. C1 ($n = 31$ [48.4%]) and C2 ($n = 33$ [%51.5]) strains had the same prevalence in our collection, and C1-M27 ($n = 22$) strains were closely related, carried a unique plasmid type (F1:A2:B20), and exhibited virotype C. Removal of 11 C2 strains with varied virotype patterns and the heterogeneous IncF type identified 22 closely related virotype E/F strains with replicon type F31/F36:A4:B1, forming what we denote as the “C2-subset.” In a global context, the C2-subset constituted a distinct cluster with international virotype E strains and harbored a genomic island, *Gl-pheU*. Association of *cnf1/hlyCABD* genes with 1 to 7 mobile genetic elements, mostly IS682/ISKpn37 combination within *Gl-pheU* was identified. The C2-subset accounted for excess resistance/virulence of subclade C2 relative to C1 strains. In addition, a conserved chromosomal IS26-mediated composite transposon (IS15DIV-ISEcp1-*bla*_{CTX-M-15}-*WbuC* cupin fold metalloprotein-Tn2-IS15DIV) was observed in the C2-subset. The local spread of the C2-subset in the hospital studied, with the carriage of higher virulence/resistance markers and a peculiar F-type plasmid, demonstrates the potential for diversification of the ST131 lineage and the emergence of subpopulations with higher survival potential to cause health care-associated outbreaks.

IMPORTANCE *Escherichia coli* sequence type 131 (ST131) is a globally dominant multi-drug-resistant clone that is commonly associated with extraintestinal infections. Specific sublineages have been shown to have emerged and spread within ST131, highlighting the complex nature of ST131 epidemiology. This study systematically compared the Iranian ST131 population to those reported from other countries and found a subpopulation harboring virotype E, a homogeneous profile of plasmid Inc-F type F31/F36:A4:B1 harboring *cnf1*/hemolysin genes on the genomic island *Gl-pheU*, and up to seven mobile genetic elements (MGEs) flanking *cnf1*/hemolysin virulence markers. The results of this study highlight the importance of MGEs for virulence gene acquisition and the formation of new subpopulations among pandemic clones such as *E. coli* ST131.

KEYWORDS ST131, whole-genome sequencing, hemolysin, genomic island, SNP phylogeny, virulence, cgMLST, Iran, cytotoxin, virulence factors

Citation Pajand O, Rahimi H, Darabi N, Roudi S, Ghassemi K, Aarestrup FM, Leekitcharoenphon P. 2021. Arrangements of mobile genetic elements among virotype E subpopulation of *Escherichia coli* sequence type 131 strains with high antimicrobial resistance and virulence gene content. *mSphere* 6:e00550-21. <https://doi.org/10.1128/mSphere.00550-21>.

Editor Maria L. Marco, University of California, Davis

Copyright © 2021 Pajand 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 Frank M. Aarestrup, fmaa@food.dtu.dk.

Received 15 June 2021

Accepted 8 August 2021

Published 25 August 2021

E*scherichia coli* sequence type 131 (ST131) is an example of successful multidrug-resistant pandemic clone among human pathogens (1, 2). ST131 is the most common clone among extraintestinal pathogenic *E. coli* (ExPEC) and shows the highest rates of fluoroquinolone (60 to 90%) and cephalosporin (40 to 80%) resistance among this pathotype (3, 4). Population genetics studies have delineated the ST131 phylogeny into three major clades (5). Clades A and B are minor subsets among ST131 clones, and clade C represents the largest clade and comprises two sublineages, C1 (*H30R*) and C2 (*H30Rx*). Clade C is uniformly fluoroquinolone resistant due to conserved replacement mutations in *gyrA* and *parC* and mostly shows cefotaxime/ceftazidime-resistant phenotypes due to producing of *bla*_{CTX-M-15} and *bla*_{CTX-M-27} enzymes. C1 subclade is mainly associated with carriage of *bla*_{CTX-M-27} or *bla*_{CTX-M-14}. In contrast, C2 subclade which is subdivided from C1 based on specific a single nucleotide polymorphism (SNP) at *fimH30*, mostly harbor *bla*_{CTX-M-15} and represents the dominant population among clade C (6). Compared to the highly conserved core genome, there is a wide plasticity in the accessory genome of ST131 resulting in appearance of different virotypes and plasmid content (7).

Based on genome sequence analyses of ST131 isolates, the major determinants of this clone have been partially characterized (8–10). The harboring of plasmids belonging to incompatibility (Inc) groups with F replicons that acquire resistance genes and spread rapidly is one of ST131 prominence. The most common pMLST types identified among multidrug resistant C1 and C2 subclades are F1:A2:B20 and F2:A1:B-, respectively (5, 11–13). Virulence gene content plays the other preponderant role in the evolution of this clone. Several studies have shown that few virulence genes are consistent among ST131 strains and there is variation in the repertoire of virulence genes (14). In this regard, the emergence of new virotypes within the C2-subclade demonstrates the evolving nature of this subpopulation, particularly with a recent report of C2/*H30Rx* strains harboring *cnf1*, *tia/hek*, and *hlyCABD* genes from Singapore (15). The importance of virulence gene acquisition is underscored by the “perfect storm” theory first proposed by Ben Zakour et al. (14), which states that this acquisition is followed by the development of antibiotic resistance in a pandemic clone such as *E. coli* ST131. The acquisition of some virulence genes such as autotransporter genes *agn-43* and *sat*, the gene cluster encoding aerobactin biosynthesis (*iucABCD*) and the gene encoding the siderophore receptor *iutA* through the acquisition of the genomic island *GI-pheV* has been shown in clade C strains (14). In our previous study of the ST131 collection, we found the occurrence of a subset under subclade C2 harboring *hlyA/cnf1* virulence genes and their association with the expression of the aminoglycoside resistance phenotype and a higher level of virulence genes (16). This motivated us to conduct a whole-genome analysis of population structure, focusing on the *cnf1/hlyCABD*-bearing subset in the context of a local hospital and global collections to further our understanding of the evolving nature of the pandemic ST131 clone. Analysis of the *cnf1/hlyCABD* carrying strains reveals that they are carried by acquisition of a horizontally transmitted genomic island, *GI-pheU*. We also describe the association of these virulence genes with up to seven MGEs on positive contigs.

RESULTS

Highest resistance rate was observed against cefotaxime. In this study, we focused on the genetic profiles of 72 ST131 *E. coli* strains isolated during March 2015 to September 2016 (19 months). Of the 72 patients with an age range of 13 to 97 years, 31 (43.1%) were male. Data regarding the sample source, MDR phenotype, and patient’s gender in terms of clade classification are presented in Table 1. Antibiotic susceptibility testing revealed the highest susceptibility rates against imipenem (97.2%), meropenem (98.6%), amikacin (88.9%), and ertapenem (87.5%). The highest resistance rates were observed against cefotaxime ($n = 67$, 93.1%). All eight clade A strains were susceptible to fluoroquinolone.

Equal distribution of subclade C1 and C2 among collected strains. Two clade A strains were characterized as EAEC. The other EAEC-associated genes, such as *aafC* and

TABLE 1 Demographic characteristics of studied ST131 isolates

Clade or subclade (total no. of isolates)	No. (%) of isolates from females	Source of isolation (n) ^a	No. (%) of isolates ^b	
			MDR	ExPEC
Clade A (8)	6 (75)	Respiratory (1), UC (7)	6 (75)	4 (50)
Subclade C1 (31)	17 (54.8)	Respiratory (5), UC (24), wound (2)	28 (90.3)	29 (93.5)
Subclade C2 (33)	18 (54.5)	UC (31), wound (2)	32 (97)	33 (100)

^aUC, urine culture; n, number of isolates.

^bMDR, multidrug resistant; ExPEC, extraintestinal pathogenic *E. coli*.

astA, were identified among nine and four strains, respectively. The resistance genes detected among the eight clade A strains were *bla*_{TEM-1B} (100%), *bla*_{CTX-M-15} (62.5%), *aadA5*, *aph6Id*, *aph3Ib*, *tet(A)* (62.5%), *aac3IId*, *tet(B)* (25%), and *bla*_{CMY-42} (12.5%). Mutations detected in two subunits of DNA-topoisomerase IV and DNA-gyrase genes in clade A strains were *parE*(I529L/S458A) and *gyrA*(S83L), respectively. The phylogeny of the 72 strains based on genomic alignment contained 7,074 SNPs totally. This recapitulated the clades A and C and showed that most strains were from clade C ($n = 64$, all fluoroquinolone resistant), specifically subclade C2 ($n = 33$), followed by subclade C1 ($n = 31$) and clade A. All clade A and C strains were *fimH41* and *fimH30*, respectively (Fig. 1).

We focused our subsequent analyses on clade C strains since they represented the most common clade.

Appearance of a subset within the C2-sublineage. Clade C strains were divided equally into the C1 ($n = 31$) and C2 ($n = 33$) subclades. C1 strains mainly belonged to C1-M27 ($n = 22$), and all except four strains harbored *bla*_{CTX-M27}. Among the nine C1-nM27 strains, eight, four, and three strains were positive for *bla*_{TEM-1B}, *bla*_{CTX-M-15}, and *bla*_{CMY-59}, respectively. As expected, there was a strong association between the carriage of *bla*_{CTX-M} (CTX-M-15 and CTX-M-27) and resistance phenotypes against ceftazidime/cefepime ($P < 0.001$). Aminoglycoside resistance genes *aph6Id* and *aph3Ib* were mostly detected among C1 (C1; $n = 20$, 64.5%, $P < 0.001$ for both genes); however, *aac6Ib-cr* and *aac3IIa* were mostly harbored by the C2-sublineage (Fig. 2a). The most common plasmid type among C1 was F1:A2: B20, while C2 strains showed varied plasmid replicon types. Genome alignment-based phylogeny showed 1,707 total SNP differences between clade C strains (Fig. 2a).

Interestingly, 22 of 33 C2 strains clustered closely in a subset. The leading branch to this subset had 100% bootstrap support, which suggest a monophyletic origin. Strains of this subset all carried *bla*_{CTX-M-15} and *bla*_{OXA-1}, and most harbored *aac6Ib-cr* ($n = 21$ [95.4%]) and *aac3IIa* ($n = 20$ [90.9%]). Here, we refer to these 22 strains as the “C2-subset.” Considering the plasmid replicon type, C2 strains were not uniform and harbored a mixed combination of plasmid types, and the C2-subset strains showed a homogeneous replicon pattern as F31 or F36:A4:B1, except for one strain, which was negative for IncF type plasmid (strain EC672).

A minimum spanning tree based on cgMLST of the C2-subset divided the 22 strains into three clusters, with allelic distances that ranged from 0 to 29 alleles. No differences were observed based on the distribution of resistance/virulence markers and pMLST replicon patterns between three identified clusters (Fig. 2b).

The C2-subset harbors higher virulence markers. Toxin (*senB*, *cnf1*, and hemolysin encoding cluster [*hlyCABD*]), invasin (*tia/hek*), adhesin (*hra*, *papK*, *papD*, *papC*, and *papGII*), protection (*kpsMTII-k5*), and contact-dependent inhibition system (*cdiB*) (Fig. 3) genes were common among the C2 subset compared to other C2 strains. Virotyping categorized 64 strains into five virotypes, with the highest prevalence of virotype C ($n = 30$ [46.8%]) and E ($n = 20$ [31.2%]). Interestingly, virotype C was the most common pattern among subclade C1, whereas C2 strains displayed virotype heterogeneity. By focusing specifically on 22 C2-subsets, all except two strains were identified as virotype E (positive for *papC*, *papGII*, *kpsMTII-k5*, *sat*, *cnf1*, and *hlyA*) and the remaining two ones

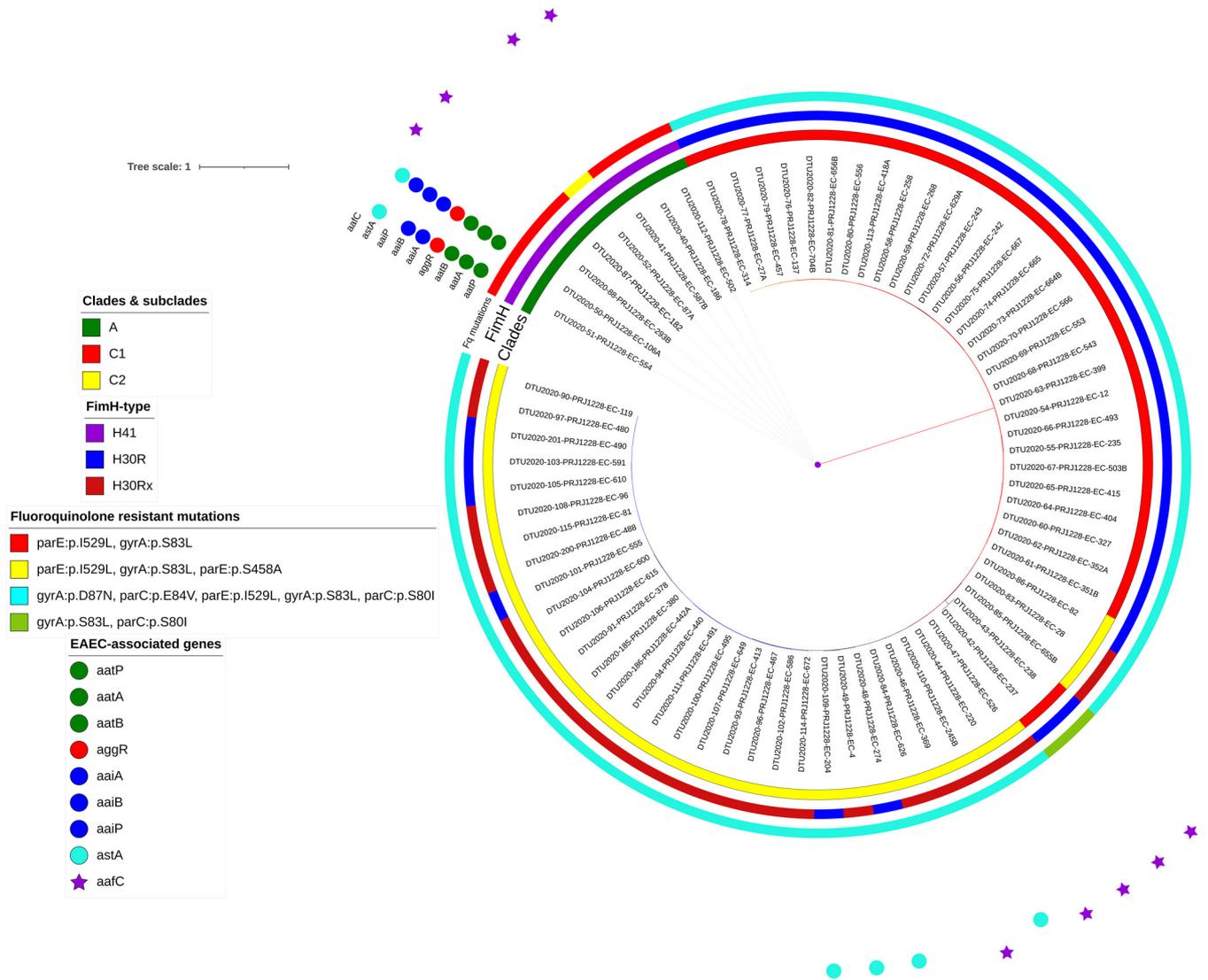


FIG 1 Phylogenetic reconstruction of $n = 72$ collected ST131 strains. Maximum-likelihood phylogeny of $n = 72$ ST131 showed two main clades and C ($n = 64$), with two common subclades in C: C1 ($n = 31$) and C2 ($n = 33$). The midpoint-rooted phylogram was constructed with PhyML from the chromosome-wide SNPs arising by mutation and visualized with iTol. Allelic profiling of *fimH*, *gyrA-parC*, clade classification, and EAEC-associated genes is represented as colored strips around the phylogenetic tree.

belonged to virotype F (*cnf1/hlyCABD* negative) and unknown (*kpsmTII-k5* negative). Of 64 clade C strains, 6 (9.3%) displayed an unknown virotype.

C1-M27 and C2-subset strains differentiated into two distinct clusters in the global context. We analyzed the 64 clade C strains from Iran in a global context using 314 publicly available clade C whole-genome sequence data present in ENA database. From the study conducted by Ludden et al. (17), 289 genomes, including strains from C1-10, C2-7, C2-8, and C2-9, Fastbaps clusters were selected. The 25 remaining genomes were selected from a collection of recently published Singaporean clade C strains isolated from bacteremia patients. A phylogenetic tree of 378 genomes (including our 64 clade C strains) showed that Iranian isolates clustered into two groups: first, C1-M27 strains that were closely related and clustered with other international *bla*_{CTX-M-27}-producing isolates, and second, a cluster including 21 of 22 C2-subset isolates that were closely grouped with 14 international genomes, all belonging to virotypes E and F. The remaining C2-subset strain that did not cluster with other *cnf1/hlyCABD*-carrying isolates was identified as unknown virotype. The 11 C2 strains were mostly diverse and distributed among different C2 clusters. We identified 20,132 SNPs by mapping of 378 genomes to

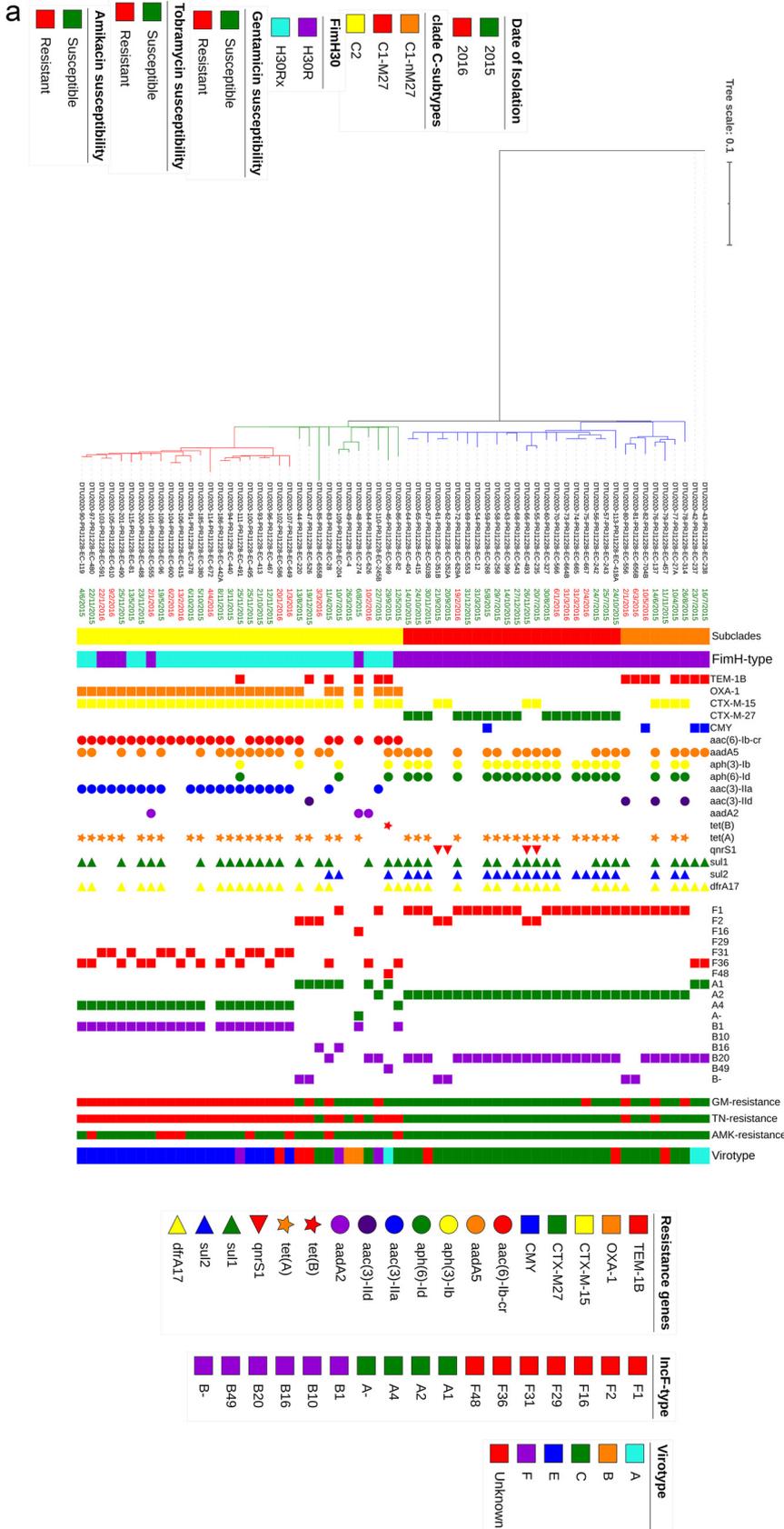


FIG 2 (a) Maximum-likelihood phylogeny of clade C strains. A phylogenetic reconstruction of 64 strains from clade C is shown. This image shows three common subclades in C, C1-nM27 ($n = 9$), C1-M27 (Continued on next page)

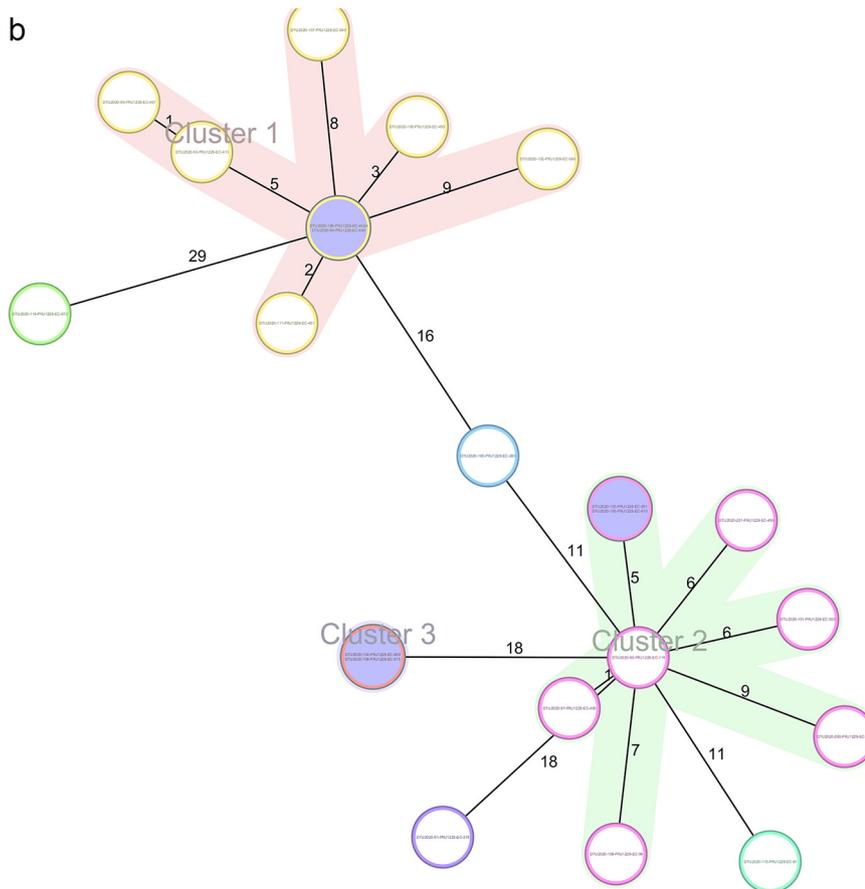


FIG 2 (Continued)

EC958. The C-subclade classification was supported by *fimH30R/Rx* allelic differences, except for 11 C2 strains which were identified as *H30R* or non-*H30*. The β -lactam resistance markers, including *bla*_{CTX-M-27} and *bla*_{OXA-17}, were exclusively detected among C1 and C2, respectively, while strains harboring *bla*_{CTX-M-15} (C2; $n = 228$ [75.2%], $P < 0.001$) were distributed among two C sublineages. Of the aminoglycoside resistance determinants (ARDs), *aac61b-cr* (C2; $n = 211$ [69.6%], $P < 0.001$) and *aac3IIa* (C2; $n = 61$ [20.1%], $P < 0.001$) were just detected among subclade C2. The other resistance markers, including *aac3IIIa* (C1; $n = 19$ [25.7%]), *aadA5* (C1; $n = 26$ [35.1%], $P = 0.002$), *aph31b* (C1; $n = 25$ [33.8%], $P < 0.001$), and *aph6Id* (C1; $n = 25$ [33.8%], $P < 0.001$), were harbored by both C subclades, with the last three were significantly detected among the C1 subclade (Fig. 4).

The *cnf1/hlyCABD* carrying strains display similar resistance/virulence/pMLST repertoire patterns. The 22 C2-subset, along with 18 C2 *cnf1/hlyCABD*-positive international isolates, was subjected to SNP phylogeny construction, and a total of 2,073 SNP differences were identified. The isolation dates for international isolates were from 2008 to

FIG 2 Legend (Continued)

($n = 22$), and C2 ($n = 33$), where the last is divided into two groups, including C2 ($n = 11$) and C2-subset ($n = 22$). The data presented in the phylogram represent the date of isolation, C-subclade classification, the *fimH* allele type, resistance genes, the plasmid Inc-F types, aminoglycoside susceptibility patterns, and virotypes. The red branches in the phylogram represents the C2-subset strains. (b) Minimum spanning tree of 22 C2-subset isolates. The genomes were analyzed using an *ad hoc E. coli* scheme based on 2,764 targets. A maximum allelic distance of ≤ 10 alleles was used to define a cluster. The isolate ID numbers are shown in the bubbles, and the allelic distances between isolates are expressed on the lines connecting them. Clusters are labeled and indicated by gray zones around the included isolates.

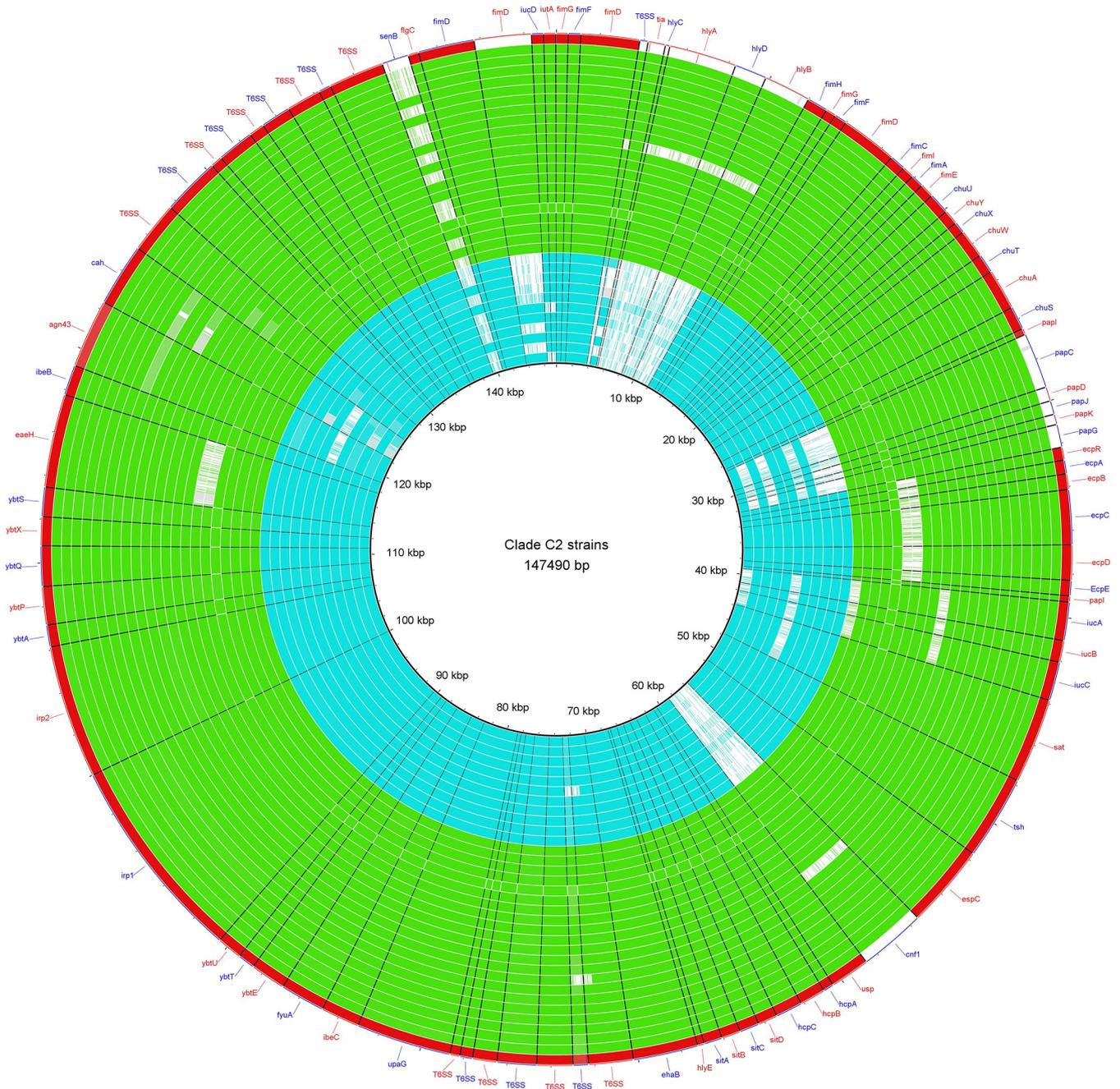


FIG 3 Virulence genes content of ST131 subclade C2. A BLAST Ring Image Generator (BRIG) image shows the presence of virulence factors among C2 (aqua rings) and C2-subset strains (green rings). The *FFN* format file of EC119 (one of the C2-subset) was downloaded from VFDB website, and open reading frames (ORFs) are annotated based on the VFDB results and used as reference strains to draw the image. The last ring (red) is EC958.

2015. All except one strain were *bla*_{CTX-M-15} positive and mostly harbored *aac6lb-cr* ($n = 36$ [90%]), *bla*_{OXA-1} ($n = 37$ [92.5%]), and *aac3IIa* and *aac3IIc* ($n = 32$ [80%]). The most common detected plasmid replicons among 40 strains were F31:A4:B1 and F36:A4:B1. Virotyping identified 35 strains as virotype E. Of the remaining five strains, one was identified as virotype F, and four were identified as *kpsMTII-k5* negative and categorized as unknown (Fig. 5).

Carriage of *cnf1/hlyCABD* genes is mediated by genomic island *pheU* in association with up to seven MGEs. BRIG analysis of mobile genetic elements (MGEs) among our C2 strains ($n = 33$) and 18 *cnf1/hlyCABD*-positive international isolates, with a recently sequenced strain C2/S65EC as the reference, identified a shared genomic island *GI-pheU*

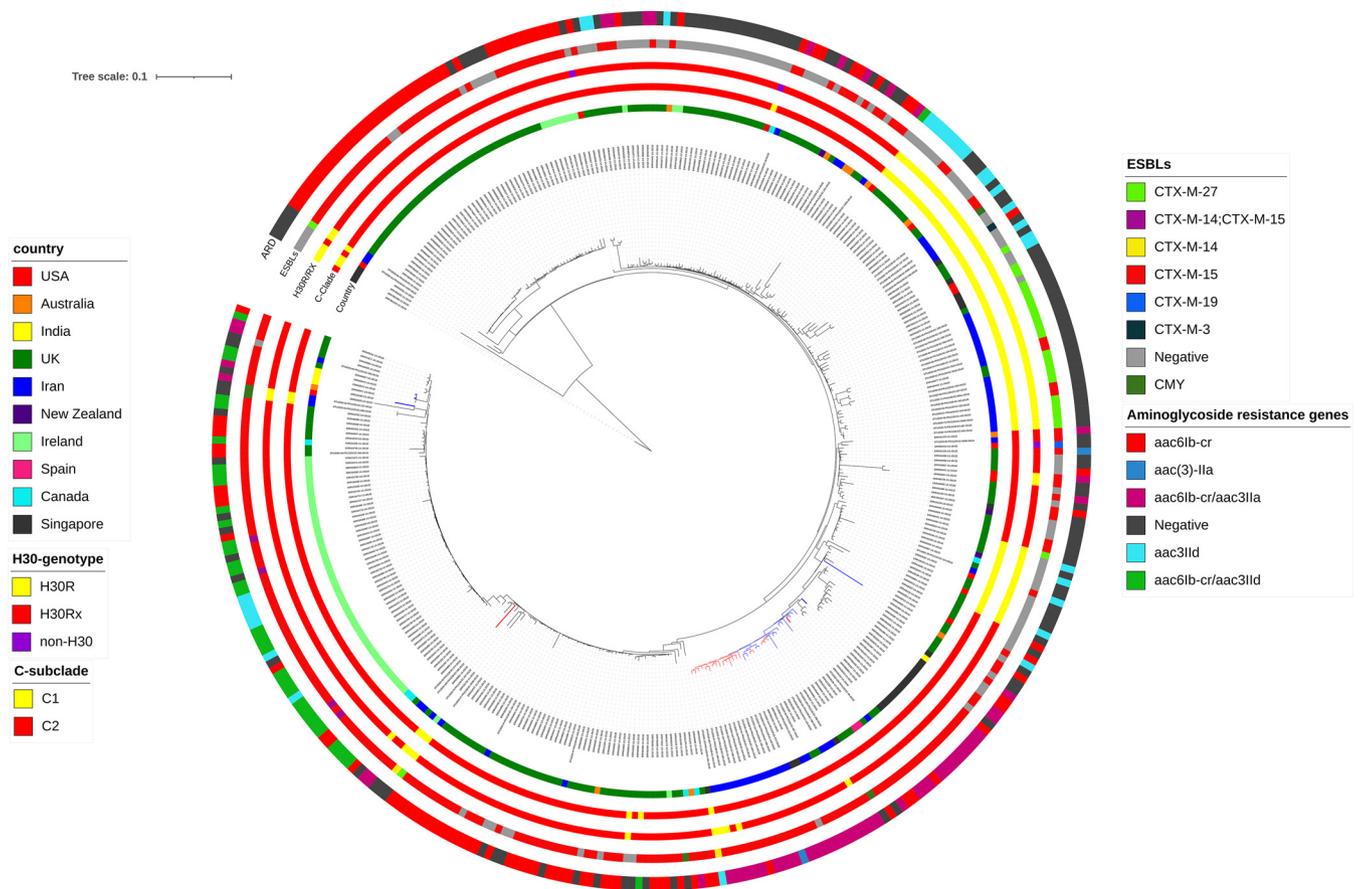


FIG 4 Maximum-likelihood phylogeny of clade C strains from the global ST131 collection. A phylogenetic reconstruction of 378 strains from clade C with EC958 as the reference strain is shown. This shows two common subclades in C, C1 ($n = 76$), and C2 ($n = 302$). The red branches indicate the C2-subset, and blue branches indicate the international *cnf1/hlyCABD*-positive genomes. Colored strips surrounding the phylogram represent the country of origin of each strain, clade classification, *fimH30* allele, ESBL genes detected alone or in combination with other ESBLs, and aminoglycoside resistance determinants (ARDs) alone or in combination with other ARDs.

among *cnf1/hlyCABD*-carrying strains, which was not detected among EC958 reference strain and other C2 isolates (Fig. 6). However, conserved structure (complete identity 100%) with *Gl-pheU* just was observed in four strains (all were international strains). Prophage *phi-8* was either partially or completely missing in all except for two strains (ERR161303 and EC274), and *phi-2* prophage was detected as intact in one international strain (ERR161303). Since S65EC was used as the reference strain in BRIG analysis and this strain is negative for *Gl-selC* compared to EC958, we could not show *Gl-selC* carriage among the studied 40 strains.

Focusing on contigs carrying *cnf1/hlyCABD* cluster and *Gl-pheU* as the reference for BRIG analysis, we identified *cnf1/hlyCABD* carriage with the association of 1 to 7 mobile genetic elements, including IS682, ISEc16, ISKpn37, IS629, ISEc10, and three composite transposons (cn_4876_IS629, cn_15494_IS629, and cn_3070) (Fig. 7). ISKpn37 was present in all except one strain which carried IS682 in association with ISEc16. The most common MGE combination patterns were ISKpn37 alone or along with ISEc16 and IS682, in the same direction of *cnf1/hlyCABD* genes (see Fig. S1a and b in the supplemental material). The genetic structure identified on all except two *cnf1/hlyCABD* carrying contigs was ISKpn37, histidine kinase, *hlyCABD/cnf1*, "helix-turn-helix domain containing protein," and IS682, of that histidine kinase and "helix-turn-helix domain containing protein" were in reversed direction. In the remaining two strains (EC467 and EC672), the truncated form of *cnf1* was identified on two contigs. The observed pattern in these two strains was in this format: ~500 bp of *cnf1*, along with *hlyCABD* operon on one contig, and the remaining ~2,500 bp on the other contig (see Fig. S1c and d). The carriage of MGE divided the C2-

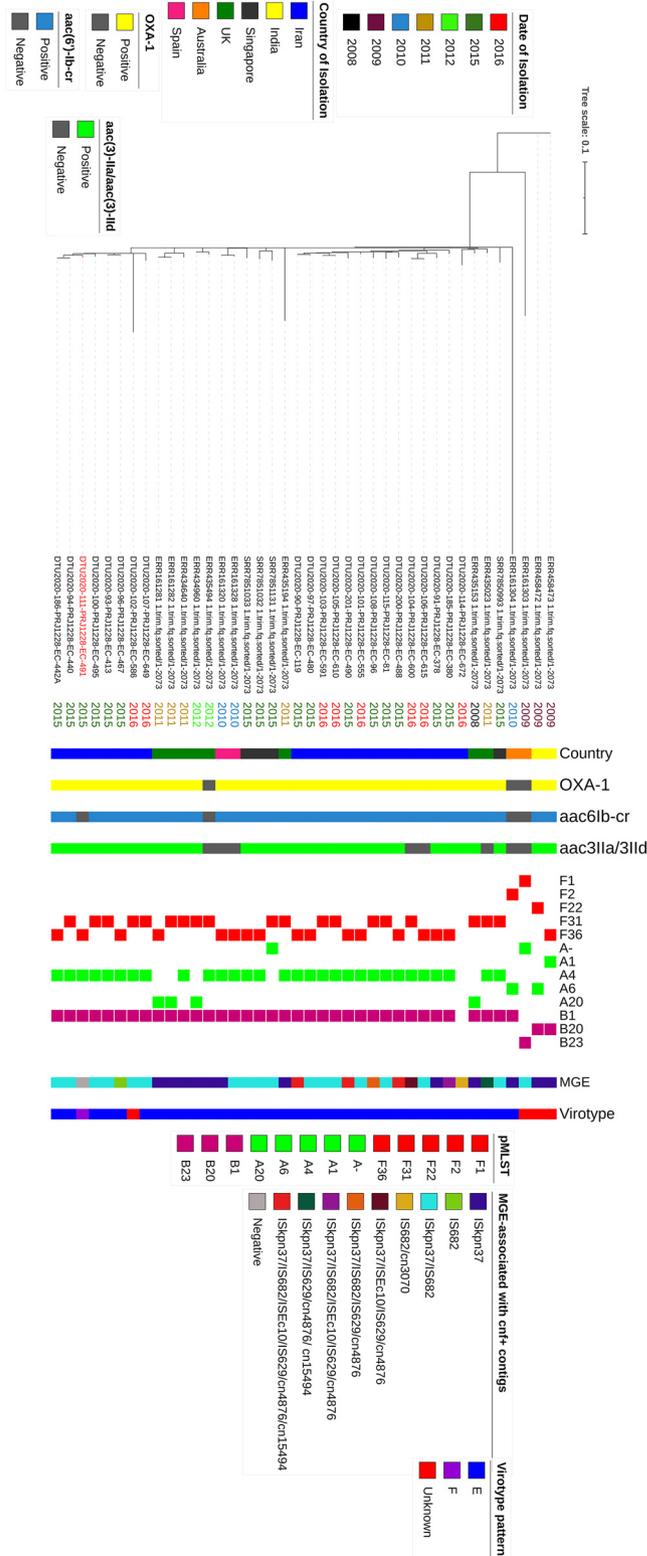


FIG 5 Maximum-likelihood phylogeny of the C2-subset and international *cnf1/hlyCABD*-carrying strains. A phylogenetic reconstruction of 40 strains from subclade C2 is shown. All except one strain were positive for *bla*_{CTX-M-15}. C2-subsets are divided into two clusters, with international isolates distributed among them. Viortype F strain (strain 491) is indicated by a red label. The data presented in the phylogram represent the date of isolation, the country of isolation, *bla*_{OXA-1}, *aac6Ib-cr*, *aac3IIa/aac3III*d, pMLST replicon types, viortypes, and MGEs associated with *cnf1/hlyCABD* genes.

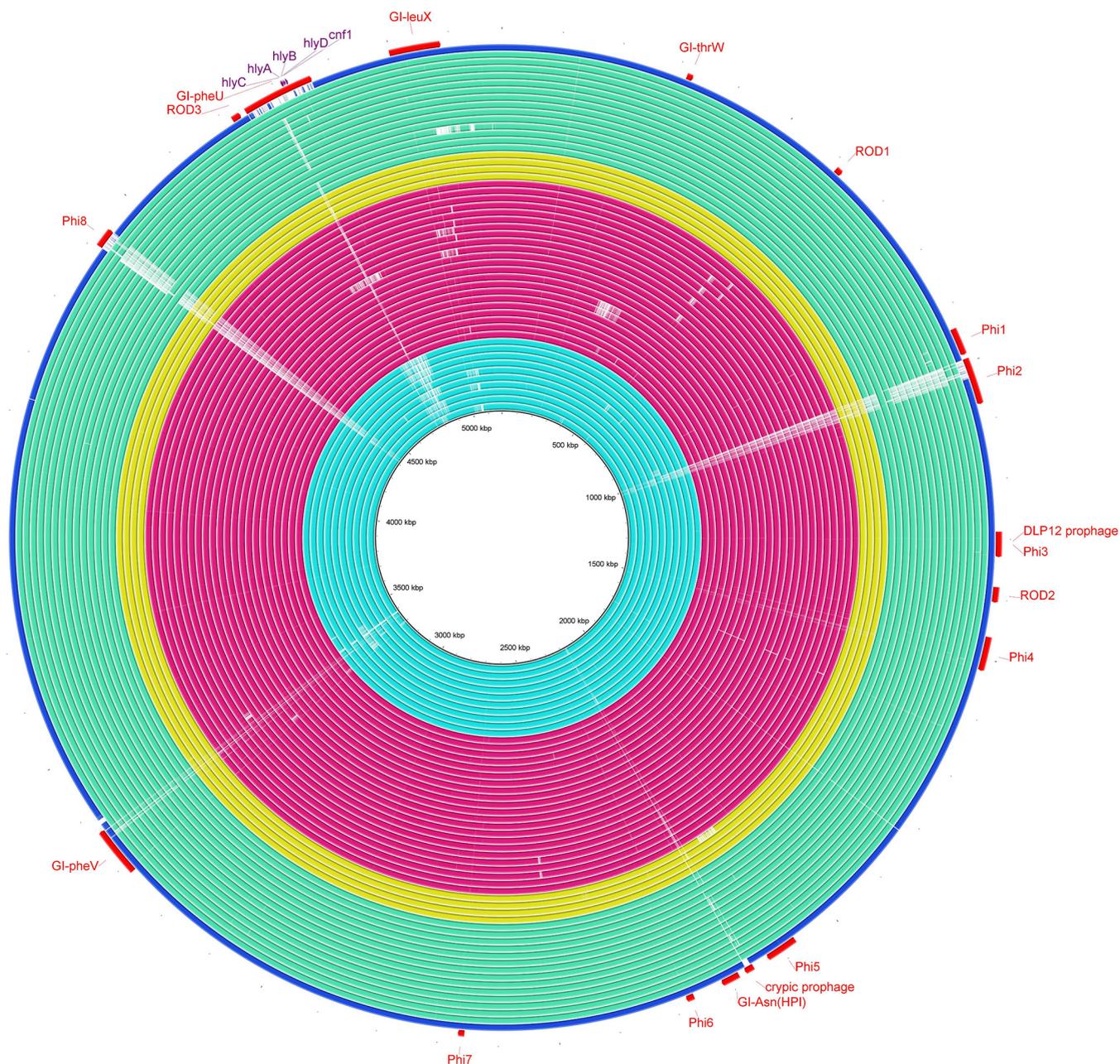


FIG 6 Genome similarities to the *Escherichia coli* ST131 S65EC genomic islands. Rings drawn by BRIG show the presence of these regions among 11 C2 (“deep sky blue” rings), 22 C2-subset (red rings), and 18 international virotype E strains (yellow and “aquamarine” rings). Colored segments indicate >70% similarity, and gray segments indicate >50% similarity by BLAST comparison between the regions of interest and each genome. The *GI-pheU* harboring *cnf1/hlyCABD* genes were detected among all except one C2-subset strain, which was negative for the mentioned virulence genes (virotype F), and also international genomes. The prophage 8 (*Phi8*) region was completely found in one C2 (EC274) and one international strain (ERR161303), and prophage 2 (*Phi2*) was partially detected among all except one international strain (ERR161303). The last ring (blue ring) indicates EC958 strain.

subset into two groups. Group 1 carried two MGEs, along with *cnf1/hlyCABD* genes, while group 2 strains harbored mixed combinations of MGEs (from one to seven elements) (Fig. 5). The remaining 18 public positive strains were associated with up to four MGEs.

Two genetic environments of *bla*_{CTX-M-15} detected in C1-M27 and the C2-subset.

To predict the cotransfer of resistance genes with our sequencing results, the assembled genomes of all 64 strains were checked. The most common cocarriage of resistance markers was *aph3Ib/aph6Id/sul2/tetA* among C1-M27 and *aac6Ib-cr/bla*_{OXA-1}/*catB3* among C2 strains; both were associated with IS6 family transposase (*IS15DIV*) (Fig. 8a and b). Also, the gentamicin/tobramycin resistance marker, *aac3IIa*, was followed by AAA-family

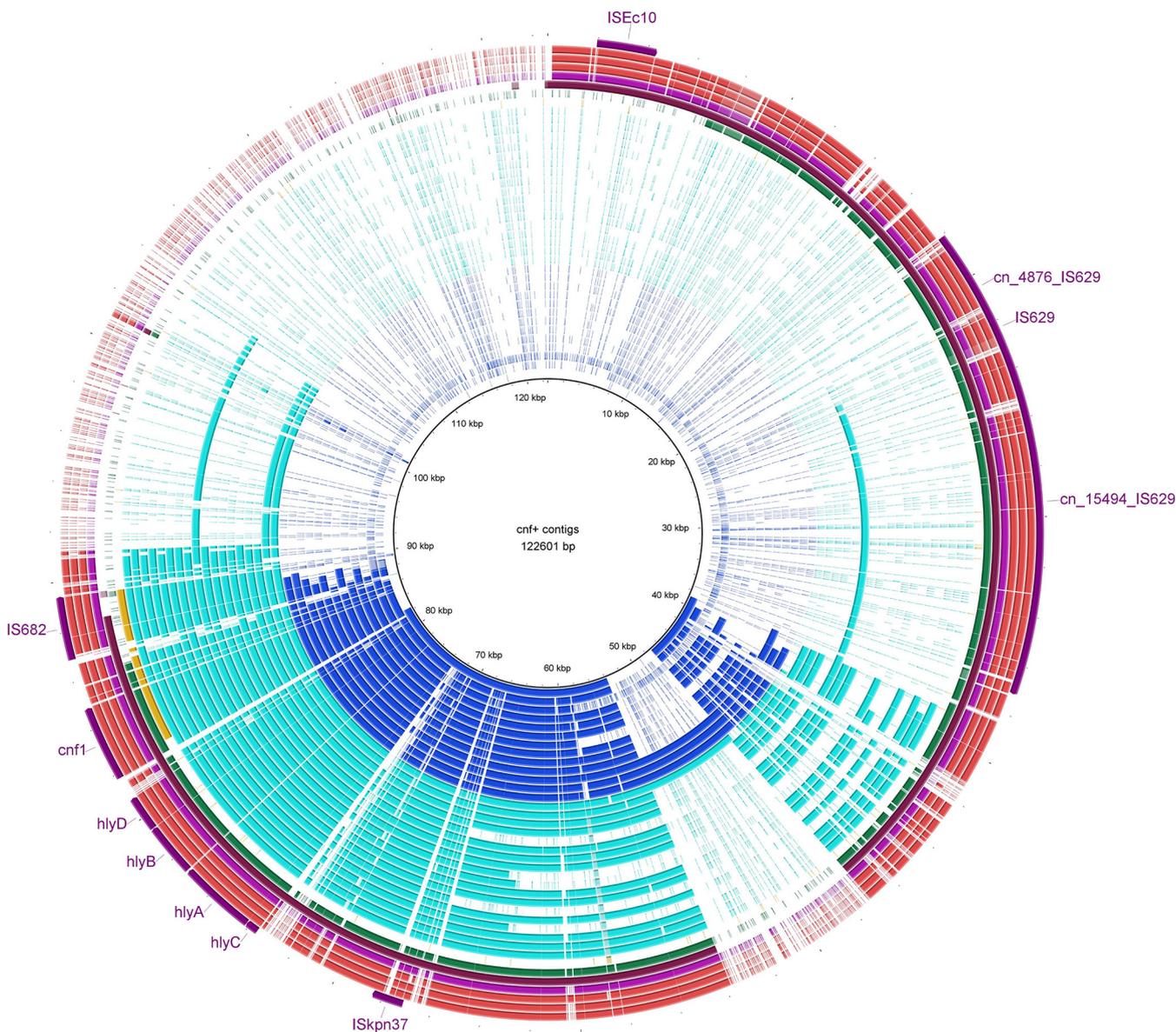


FIG 7 Mobile genetic elements within contigs carrying *cnf1/hlyCABD*. A BLAST Ring Image Generator (BRIG) image shows the presence of the MGEs reported in *GI-pheU* (122,601 bp) in association with *cnf1/hlyCABD* among our data set. Colored segments indicate >70% similarity, and gray segments indicate >50% similarity by BLAST comparison between the regions of interest and each contig. Blue rings indicate contigs harboring one MGE (*ISkpn37*), aqua rings indicate contigs harboring two MGEs (*IS682* and *ISkpn37*), yellow rings indicate strain EC672 with a truncated *cnf1*, and the last six rings are strains harboring four to six MGEs in association with *cnf1/hlyCABD*.

ATPase and a large insertion sequence *ISkpn11*, which were detected among all C2 strains (Fig. 8c). Considering the *bla*_{CTX-M-15}-carrying contigs, two different patterns were observed. First, the *bla*_{CTX-M-15} was followed by left-truncated *ISEcp1* and preceded by cupin fold metalloprotein *WbuC*, a large Tn2, IS3 transposase, *qnrS1*, *ISkpn19*, and plasmid-related genes, including recombinase and addiction genes in three of the four C1-M27 strains. The *WbuC*, *ISkpn19*, IS3 transposase, and plasmid-related genes were located in the reverse direction (Fig. 8d).

The other genetic structure of *bla*_{CTX-M-15} gene observed on small contigs mostly among the C2-subset was *ISEcp1* upstream of and in the same orientation as *bla*_{CTX-M-15}, the reversed *wbuC* cupin fold metalloprotein, and a large Tn2 transposon (Fig. 8e). In both identified genetic structures, *bla*_{CTX-M-15} was located between *IS15DI* (IS6 family transposase), reversed *WbuC*, and a large Tn2.

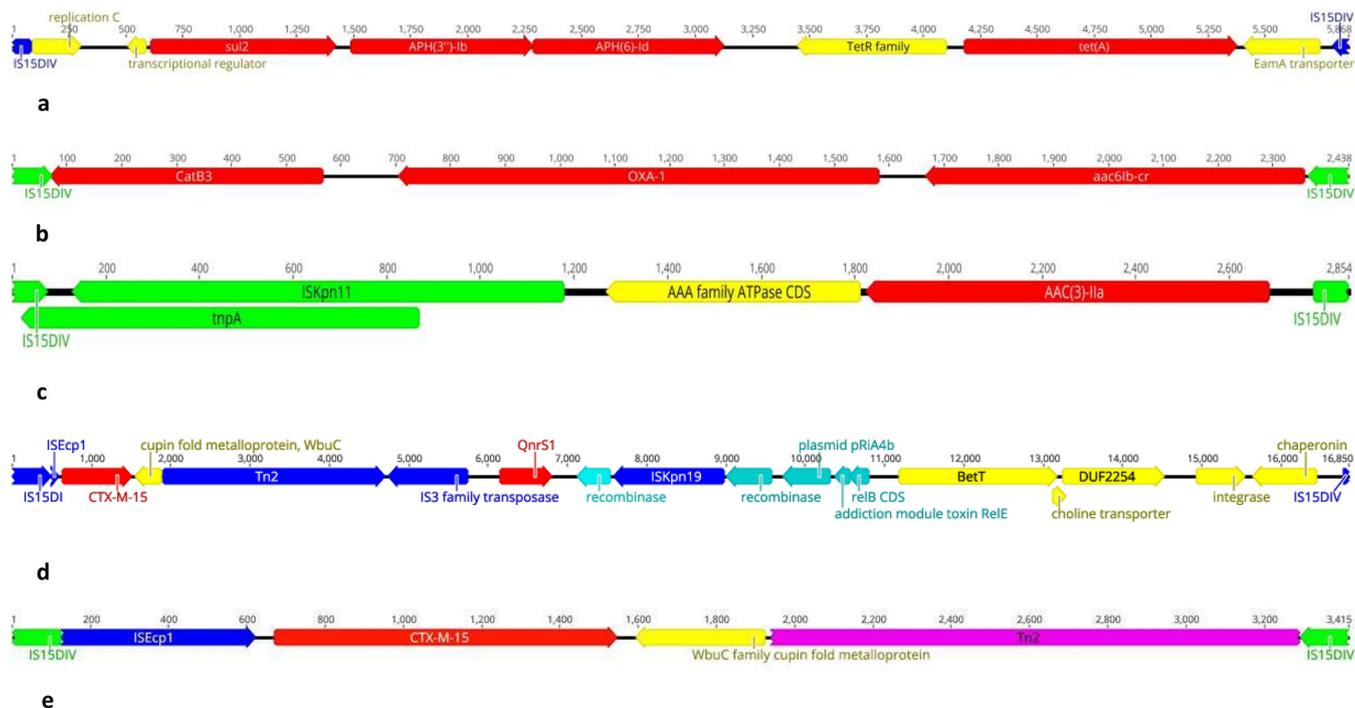


FIG 8 Genetic environment of the most common resistance genes among study isolates. (a to c) Contig harboring *IS15DIV-sul2-aph3lb-aph6ld-tetA-IS15DIV* (a), *IS15DIV-catB3-bla_{OXA-1}-aac6lb-cr-IS15DIV* (b), and *IS15DIV-ISKpn11-aac3IIa-IS15DIV* (c). (d and e) Two forms of the *bla_{CTX-M-15}* genetic environment. Annotation has been done by Geneious Prime 2021.1.1.

DISCUSSION

Along with our previous findings which have indicated a dynamic change toward the emergence of a subset harboring *cnf1/hlyA* virulence genes with aminoglycoside resistance phenotypes (16), we performed whole-genome sequencing of all ST131 isolates submitted from our health care center over 19 months. We performed sequence variation analysis on the globally dominant clade C isolates focusing on the *cnf1/hlyCABD*-positive ones. Using bioinformatics analyses, we examined the relatedness of positive strains and MGE association with the carriage of these virulence markers. Finally, we also applied genome annotation to identify the genetic environments of contigs carrying *cnf1/hlyCABD* and the resistance gene *bla_{CTX-M-15}*.

The ST131 population in Semnan province was dominated by clade C, which is similar to results from a previous large global survey (17). Although C2 strains were found to be the most resistant subclade in our collection and were associated with *bla_{CTX-M-15}/aac6lb-cr/aac3IIa* and carriage of higher virulence markers, their prevalence was equal to the C1-subclade. The priority of the C1-subclade and specifically the C1-M27 subclade could be explained by these features: lower SNP differences and consequently higher clonality compared to C2 or even clade A strains, the great homogeneity of its plasmid content, and the virulence patterns (mostly virotype C) (18). However, the clonal displacement of different sublineages among ST131 population has been reported recently from Ireland in that the C1/CTX-M-14-producing subpopulation was displaced by the expansion of a genetically distinct C2 lineage referred to as C2-8 harboring chromosomally located CTX-M-15 in the *mppA* gene (17). Nevertheless, C1-M27 still has been reported as the dominant ST131 population from different countries such as Germany, France, and Japan (6, 19).

In the context of global ST131 genomes we found that two groups of Iranian strains, C1-M27 and C2-subset, which were equal in prevalence ($n = 22$ [34.3%]), have constituted distinct clusters. C1-M27 and C2 strains display two striking common features: first, producing cefotaxime/ceftazidime CTX-M hydrolyzing enzyme and, second,

carriage of IncF-type plasmids encoding different resistance/virulence markers. While IncF-type plasmid among C2 was not as homogeneous as the C1-M27 subclone overall, focusing on the C2-subset revealed a unique pattern, namely, the F31/F36:A4:B1 replicon type. Again, most of the C2-subset displayed the same virotype (virotype E) showed a monophyletic origin and a common pattern of virulence/resistance markers. This C2-subset was also responsible for association of C2 population with higher virulence content and aminoglycoside resistance phenotype; nonetheless, this subpopulation seemed to be similar to other C2 strains overall in terms of its virulence/resistance repertoire. Therefore, it seems that possessing of virulence factors such as toxins (*cnf1*, hemolysin), invasins (*ibeB*, *tia/hek*), and adhesins (*hra*, *papC/GII*) likely contributes to the virulence potential of the C2-subset, and their overrepresentation in this subpopulation is related to the monophyletic nature of this subset.

The carriage of genomic island *Gl-pheU* demonstrated among all *cnf1/hlyCABD*-positive strains (virotype E) and was not detected among EC958 or other C2 strains. Nonetheless, *Gl-pheU* also carried other virulence markers, such as F17 fimbriae, *papE*, or *papA_F43*, which were absent among our strains. IS*Kpn37* and IS682 were the most common detected MGEs flanking the *cnf1/hlyCABD* genes. However, two strains were found to display different patterns; in both, the *cnf1* gene was truncated and found on two contigs. Therefore, it seems that while *Gl-pheU* acquisition plays a key role in the transfer of *cnf1/hlyCABD* genes among C2 strains, insertion sequences such as IS682 are the second determinant in the transposition of *cnf1* inside the chromosome.

Diverse *bla*_{CTX-M-15} genetic environments have been reported among *E. coli* isolated all over the world (20). Structures containing the intact *ISEcp1*, followed by *bla*_{CTX-M-15} and reversed *WbuC* cupin fold metalloprotein, have been identified in many different species (21). In our strain collection, two forms of CTX-M-15 genetic structure were observed, with the shared features of *bla*_{CTX-M-15}, reversed *WbuC* metalloprotein, and Tn2. The unit, combined with the bounded two copies of IS15*DNV* (an IS6 family member with a sequence and function similar to that of IS26), constitutes a composite transposon that is able to transfer by forming a circular molecule (22). The carriage of an identical IS6-mediated composite transposon has been reported mainly among *E. coli* plasmids (23). The other most common MGE in our strains, *ISEcp1* (member of the IS1380 family), has been shown to display two functions in association with *bla*_{CTX-M-15}. First, it encodes a transposase that facilitates CTX-M-15 mobilization, and second, it promotes the weakly expression state of this ESBL gene by playing as a promoter (24). A histidine residue as a metal-binding ligand was identified in the *WbuC* family protein and has been reported to relate to O-antigen biosynthesis (20). The other RAST (Rapid Annotation using Subsystem Technology) annotation of the *WbuC* gene at this position was tryptophane synthase, due to the similar sequence protein that has been reported in *Klebsiella* spp. (20). Varied roles of *WbuC* have been attributed to this metalloprotein among different species; however, its specific function here with *bla*_{CTX-M-15} needs to be studied further. Another, less-reported *bla*_{CTX-M-15} genetic environment that was identified among three C1-M27 strains was IS*Kpn19-qnrS1-IS3* transposase-Tn2-*WbuC-bla*_{CTX-M-15}-IS15*DNV*. Similar structure units linked with IS*Kpn19*, instead of IS26, were seen in the chromosomes and plasmids of different bacterial species from China and other countries, suggesting that both IS26 and IS*Kpn19* contribute to the circulation of a *qnrS1*- and *bla*_{CTX-M-15}-carrying unit between chromosomes and plasmids of different bacterial species (23).

In summary, our study is notable for examining ST131 clone in an undersampled geographical region. This allowed us to identify the local epidemiology of this high-risk clone. Our findings highlight a subpopulation harboring *cnf1/hlyCABD* genes combined with up to seven MGEs (mostly IS682/IS*Kpn37*) within the horizontally acquired genomic island *Gl-pheU*. The diversity observed among ST131 lineages and resistance/virulence markers emphasizes the role of surveillance strategies to identify ST131 subpopulations, plasmids, and MGEs. Focused attention on successful subpopulations

could be helpful to identify the dynamics of emergence and their spread and control the epidemic resistance of *E. coli*.

MATERIALS AND METHODS

Bacterial collection, sequencing, and antibiotic susceptibility testing. A total of 72 *E. coli* ST131 isolates from a single tertiary referral hospital (Kosar) covering a >120,000 population of Semnan province-Iran were collected. Isolates were collected as part of standard care for admitted patients from extraintestinal specimens, including urine, blood, wound, and respiratory samples. The demographic data are shown in Table 1.

To extract genomic DNA, an Invitrogen Easy-DNA kit was used, and the DNA concentrations were determined using a Qubit dsDNA BR assay kit (Invitrogen). According to the Illumina protocol, library preparation was conducted, and sequencing was performed on Illumina NextSeq and MiSeq platforms using 150-bp paired-end reads. SPAdes was used to *de novo* assemble raw reads (25).

Antibiotic susceptibility testing was performed by using a standard disk diffusion method for the 16 antimicrobial agents. The results were interpreted according to Clinical and Laboratory Standard Institute (CLSI) guidelines (26). Isolates with intermediate or resistant phenotypes to a given antimicrobial agent were considered nonsusceptible. Multidrug resistance (MDR) was defined as resistance to at least one representative of three or more antimicrobial classes (27). An extended-spectrum beta-lactamase phenotype was confirmed by using a phenotypic combined disk test according to CLSI recommendations (26).

Genome sequence analysis. Assembled sequences were analyzed using *in silico* bioinformatics tools to confirm the serotype, sequence type, plasmid replicons, acquired antimicrobial resistance genes/fluoroquinolone resistance-associated mutations, virulence genes, plasmid MLST (pMLST) using the following pipelines: SeroTypeFinder 2.0 (12), MLST 2.0 (28), PlasmidFinder 2.1 (29), ResFinder 4.1 (30), VirulenceFinder 2.0 (31), and pMLST 2.0 (29) available from CGE. For virulence genes, the VFDB database was also used as the web interface (32). The raw-FastQ reads ($n = 289$ genomes) from a recently published comprehensive study were downloaded from European Nucleotide Archive (ENA) (17). Furthermore, 25 ST131 raw reads recently reported from Singapore were also included (15). In total, 314 ST131 short read data sets were subjected to the same pipeline mentioned above for *de novo* assembly, and the generated assemblies were processed using the databases described above. The genome annotation of some related resistance/virulence harboring contigs was performed with RAST, and .gbk and or Excel files format were downloaded. The association of resistance/virulence markers with MGEs was investigated by MGEFinder 1.0 (available from CGE) (33). The genetic environments of contigs harboring the *cnf/hly* virulence and *bla*_{CTX-M-15} resistance genes were graphically annotated by Geneious Prime 2021.1.1, and insertion sequence positions were manually curated using ISfinder BLAST results (34). BLAST Ring Image Generator (BRIG) was used to determine the similarity of genomes to ST131 genomic islands (35).

Reference based analysis was used to create the phylogenetic tree using reference strain EC958 (36). The paired-end reads were mapped to the reference genome using Burrows-Wheeler Aligner (BWA), version 0.7.2 (37). SNP calling was done using the mpileup module in SAMTools version 0.1.18. The SNP selection was based on the following criteria: (i) a distance of minimum 15 bp between each SNP (pruning), (ii) an average depth of minimum 10%, (iii) a quality of mapping of >30, (iv) an SNP quality of >20, and (v) the exclusion of all indels. Qualified SNPs from each genome were concatenated to a single alignment corresponding to the position of the reference genome. The parsimony tree construction based on PhyML was used for concatenated sequences (38) with an HKY85 substitution model and 100 bootstrap replicates. The generated phylogenetic tree was visualized and labeled with the iTOL tool. Based on cgMLST allelic mismatch between strains, a minimum spanning tree (MST) was designed by uploading the assembled genomes to Ridom SeqSphere (Ridom GmbH, Munster, Germany). Clonal clustering was defined based on a maximum of 10 allelic differences.

Clades A, C1, and C2 were identified based on matching strain identifiers with the previous report by Matsumura et al. (39). H30Rx was defined based on the specific single nucleotide variant *ybbW*(G723A), as identified by BLASTn (40). Isolates were defined as ExPEC if they were positive for ≥ 2 for any *pap*, any *sfa* or *foc*, any *afa* or *dra*, or any *kps* or *iutA* (41). Virotypes were determined based on the virulence genes scheme (42). Isolates with ≥ 1 of the enteroaggregative *E. coli* (EAEC)-associated genes *aggR*, *aatA*, *aaIC*, and *aaP* were molecularly qualified as EAEC (43).

Statistical analysis. The IBM-SPSS statistics 22 was used for all statistical analysis. Bivariate analyses were performed by using a chi-square test for categorical variables. All *P* values were two-sided.

Data availability. Raw sequence data have been submitted to the European Nucleotide Archive under study accession number [PRJEB46895](https://www.ebi.ac.uk/ena/browser/view/PRJEB46895).

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

DATA SET S1, XLSX file, 0.03 MB.

FIG S1, TIF file, 1.5 MB.

ACKNOWLEDGMENTS

We gratefully acknowledge Semnan University of Medical Sciences for a 1-year scholarship granted to Omid Pajand, as well as the staff of the Center for Genomic Epidemiology (CGE) section of DTU-Food.

REFERENCES

- Mathers AJ, Peirano G, Pitout JD. 2015. The role of epidemic resistance plasmids and international high-risk clones in the spread of multidrug-resistant *Enterobacteriaceae*. *Clin Microbiol Rev* 28:565–591. <https://doi.org/10.1128/CMR.00116-14>.
- Peirano G, Pitout JD. 2010. Molecular epidemiology of *Escherichia coli* producing CTX-M beta-lactamases: the worldwide emergence of clone ST131 O25:H4. *Int J Antimicrob Agents* 35:316–321. <https://doi.org/10.1016/j.ijantimicag.2009.11.003>.
- Manges AR, Geum HM, Guo A, Edens TJ, Fibke CD, Pitout J. 2019. Global extraintestinal pathogenic *Escherichia coli* (ExPEC) lineages. *Clin Microbiol Rev* 32:e00135–18. <https://doi.org/10.1128/CMR.00135-18>.
- Peirano G, Pitout JDD. 2019. Extended-spectrum β -lactamase-producing *Enterobacteriaceae*: update on molecular epidemiology and treatment options. *Drugs* 79:1529–1541. <https://doi.org/10.1007/s40265-019-01180-3>.
- Pitout JD, DeVinney R. 2017. *Escherichia coli* ST131: a multidrug-resistant clone primed for global domination. *F1000 Res* 6:195. <https://doi.org/10.12688/f1000research.10609.1>.
- Matsumura Y, Pitout JD, Gomi R, Matsuda T, Noguchi T, Yamamoto M, Peirano G, DeVinney R, Bradford PA, Motyl MR, Tanaka M, Nagao M, Takakura S, Ichiyama S. 2016. Global *Escherichia coli* sequence type 131 clade with *bla*_{CTX-M-27} gene. *Emerg Infect Dis* 22:1900–1907. <https://doi.org/10.3201/eid2211.160519>.
- Petty NK, Ben Zakour NL, Stanton-Cook M, Skippington E, Totsika M, Forde BM, Phan MD, Gomes Moriel D, Peters KM, Davies M, Rogers BA, Dougan G, Rodriguez-Baño J, Pascual A, Pitout JD, Upton M, Paterson DL, Walsh TR, Schembri MA, Beatson SA. 2014. Global dissemination of a multidrug-resistant *Escherichia coli* clone. *Proc Natl Acad Sci U S A* 111:5694–5699. <https://doi.org/10.1073/pnas.1322678111>.
- Schembri MA, Zakour NLB, Phan M-D, Forde BM, Stanton-Cook M, Beatson SA. 2015. Molecular characterization of the multidrug resistant *Escherichia coli* ST131 clone. *Pathogens* 4:422–430. <https://doi.org/10.3390/pathogens4030422>.
- Miyoshi-Akiyama T, Sherchan JB, Doi Y, Nagamatsu M, Sherchand JB, Tandukar S, Ohmagari N, Kirikae T, Ohara H, Hayakawa K. 2016. Comparative genome analysis of extended-spectrum-lactamase-producing *Escherichia coli* sequence type 131 strains from Nepal and Japan. *mSphere* 1:e00289–16. <https://doi.org/10.1128/mSphere.00289-16>.
- Kallonen T, Brodrick HJ, Harris SR, Corander J, Brown NM, Martin V, Peacock SJ, Parkhill J. 2017. Systematic longitudinal survey of invasive *Escherichia coli* in England demonstrates a stable population structure only transiently disturbed by the emergence of ST131. *Genome Res* 27:1437–1449. <https://doi.org/10.1101/gr.216606.116>.
- Johnson TJ, Danzeisen JL, Youmans B, Case K, Llop K, Munoz-Aguayo J, Flores-Figueroa C, Aziz M, Stoesser N, Sokurenko E, Price LB, Johnson JR. 2016. Separate F-Type plasmids have shaped the evolution of the H30 subclone of *Escherichia coli* sequence type 131. *mSphere* 4:e00121-16. <https://doi.org/10.1128/mSphere.00121-16>.
- Joensen KG, Tetzschner AMM, Iguchi A, Aarestrup FM, Scheut F. 2015. Rapid and easy in silico serotyping of *Escherichia coli* isolates by use of whole-genome sequencing data. *J Clin Microbiol* 53:2410–2426. <https://doi.org/10.1128/JCM.00008-15>.
- Jamborova I, Johnston BD, Papousek I, Kachlikova K, Micekova L, Clabots C, Skalova A, Chudejova K, Dolejska M, Literak I, Johnson JR. 2018. Extensive genetic commonality among wildlife, wastewater, community, and nosocomial isolates of *Escherichia coli* sequence type 131 (H30R1 and H30Rx subclones) that carry *bla*_{CTX-M-27} or *bla*_{CTX-M-15}. *Antimicrob Agents Chemother* 62:e00519–18. <https://doi.org/10.1128/AAC.00519-18>.
- Ben Zakour NL, Alsheikh-Hussain AS, Ashcroft MM, Khanh Nhu NT, Roberts LW, Stanton-Cook M, Schembri MA, Beatson SA. 2016. Sequential acquisition of virulence and fluoroquinolone resistance has shaped the evolution of *Escherichia coli* ST131. *mBio* 7:e00347-16. <https://doi.org/10.1128/mBio.00347-16>.
- Chen SL, Ding Y, Apisarnthanarak A, Kalimuddin S, Archuleta S, Omar SFS, De PP, Koh TH, Chew KL, Atiya N, Suwantarant N, Velayuthan RD, Wong JGX, Lye DC. 2019. The higher prevalence of extended spectrum beta-lactamases among *Escherichia coli* ST131 in Southeast Asia is driven by expansion of a single, locally prevalent subclone. *Sci Rep* 9:13245. <https://doi.org/10.1038/s41598-019-49467-5>.
- Hojabri Z, Darabi N, Mirmohammadkhani M, Rahimi H, Hemmati R, Saeedi Z, Roustaei K, Leekitchareonphon P, Pajand O, Aarestrup FM. 2020. Expansion of a subset within the C2 subclade of *Escherichia coli* sequence type 131 (ST131) is driving the increasing rates of aminoglycoside resistance. *Open Forum Infect Dis* 7:ofaa410. <https://doi.org/10.1093/ofid/ofaa410>.
- Ludden C, Decano AG, Jamroz D, Pickard D, Morris D, Parkhill J, Peacock SJ, Cormican M, Downing T. 2020. Genomic surveillance of *Escherichia coli* ST131 identifies local expansion and serial replacement of subclones. *Microb Genom* 6:e000352.
- Birgy A, Levy C, Nicolas-Chanoine MH, Cointe A, Hobson CA, Magnan M, Bechet S, Bidet P, Cohen R, Bonacorsi S. 2019. Independent host factors and bacterial genetic determinants of the emergence and dominance of *Escherichia coli* sequence type 131 CTX-M-27 in a community pediatric cohort study. *Antimicrob Agents Chemother* 63:e00382-19. <https://doi.org/10.1128/AAC.00382-19>.
- Merino I, Hernández-García M, Turrientes M-C, Pérez-Viso B, López-Fresneña N, Díaz-Agero C, Maechler F, Fankhauser-Rodríguez C, Kola A, Schrenzel J, Harbarth S, Bonten M, Gastmeier P, Canton R, Ruiz-Garbajosa P, R-GNOSIS Study Group. 2018. Emergence of ESBL-producing *Escherichia coli* ST131-C1-M27 clade colonizing patients in Europe. *J Antimicrob Chemother* 73:2973–2980. <https://doi.org/10.1093/jac/dky296>.
- Zhong Y, Guo S, Seow KLG, Ming GOH, Schlundt J. 2021. Characterization of extended-spectrum beta-lactamase-producing *Escherichia coli* isolates from Jurong Lake, Singapore, with whole-genome sequencing. *Int J Environ Res Public Health* 18:937. <https://doi.org/10.3390/ijerph18030937>.
- Dhanji H, Patel R, Wall R, Doumith M, Patel B, Hope R, Livermore DM, Woodford N. 2011. Variation in the genetic environments of *bla*_{CTX-M-15} in *Escherichia coli* from the faeces of travelers returning to the United Kingdom. *J Antimicrob Chemother* 66:1005–1012. <https://doi.org/10.1093/jac/dkr041>.
- Harmer CJ, Moran RA, Hall RM. 2014. Movement of IS26-associated antibiotic resistance genes occurs via a translocatable unit that includes a single IS26 and preferentially inserts adjacent to another IS26. *mBio* 5:e01801-14. <https://doi.org/10.1128/mBio.01801-14>.
- Li L, Olsen RH, Song A, Xiao J, Wang C, Meng H, Shi L. 2021. First report of a foodborne *Salmonella enterica* serovar Gloucester (4i:l:w) ST34 strain harboring *bla*_{CTX-M-55} and *qnrS* genes located in IS26-mediated composite transposon. *Front Microbiol* 12:646101. <https://doi.org/10.3389/fmicb.2021.646101>.
- Gourbeyre E, Siguier P, Chandler M. 2014. Bacterial insertion sequences: their genomic impact and diversity. *FEMS Microbiol Rev* 38:865–891. <https://doi.org/10.1111/1574-6976.12067>.
- Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Pribelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 19:455–477. <https://doi.org/10.1089/cmb.2012.0021>.
- Clinical and Laboratory Standards Institute. 2016. Performance standards for antimicrobial susceptibility testing; 20th informational supplement M100S-S26. CLSI, Wayne, PA.
- Pajand O, Darabi N, Arab M, Ghorbani R, Bameri Z, Ebrahimi A, Hojabri Z. 2020. The emergence of the hypervirulent *Klebsiella pneumoniae* (hvKp) strains among circulating clonal complex 147 (CC147) harboring *bla*_{NDM/OXA-48} carbapenemases in a tertiary care center of Iran. *Ann Clin Microbiol Antimicrob* 19:12. <https://doi.org/10.1186/s12941-020-00349-z>.
- Larsen MV, Cosentino S, Rasmussen S, Friis C, Hasman H, Marvig RL, Jelsbak L, Sicheritz-Pontén T, Ussery DW, Aarestrup FM, Lund O. 2012. Multilocus sequence typing of total-genome-sequenced bacteria. *J Clin Microbiol* 50:1355–1361. <https://doi.org/10.1128/JCM.06094-11>.

29. 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>.
30. Clausen PTLC, Aarestrup FM, Lund O. 2018. Rapid and precise alignment of raw reads against redundant databases with KMA. *BMC Bioinformatics* 19:307. <https://doi.org/10.1186/s12859-018-2336-6>.
31. Malberg-Tetzschner AM, Johnson JR, Johnston BD, Lund O. 2020. *In silico* genotyping of *Escherichia coli* isolates for extraintestinal virulence genes by use of whole-genome sequencing data. *J Clin Microbiol* 58:e01269-20. <https://doi.org/10.1128/JCM.01269-20>.
32. Chen L, Zheng D, Liu B, Yang J, Jin Q. 2016. VFDB 2016: hierarchical and refined dataset for big data analysis: 10 years on. *Nucleic Acids Res* 44:D694–D697. <https://doi.org/10.1093/nar/gkv1239>.
33. Johansson MHK, Bortolaia V, Tansirichaiya S, Aarestrup FM, Roberts AP, Petersen TN. 2021. Detection of mobile genetic elements associated with antibiotic resistance in *Salmonella enterica* using a newly developed web tool: MobileElementFinder. *J Antimicrob Chemother* 76:101–109. <https://doi.org/10.1093/jac/dkaa390>.
34. Siguier P, Perochon J, Lestrade L, Mahillon J, Chandler M. 2006. ISfinder: the reference centre for bacterial insertion sequences. *Nucleic Acids Res* 34:D32–D36. <https://doi.org/10.1093/nar/gkj014>.
35. Ben Zakour NL, Stanton-Cook M, Phan M-D, Totsika M, Peters KM, Chan KG, Schembri MA, Upton M, Beatson SA. 2014. The complete genome sequence of *Escherichia coli* EC958: a high-quality reference sequence for the globally disseminated multidrug-resistant *E. coli* O25b:H4–ST131 clone. *PLoS One* 9:e104400. <https://doi.org/10.1371/journal.pone.0104400>.
36. Kaas RS, Leekitcharoenphon P, Aarestrup FM, Lund O. 2014. Solving the problem of comparing whole bacterial genomes across different sequencing platforms. *PLoS One* 9:e104984. <https://doi.org/10.1371/journal.pone.0104984>.
37. Li H, Durbin R. 2009. Fast and accurate short read alignment with BurrowsWheeler transform. *Bioinformatics* 25:1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>.
38. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* 59:307–321. <https://doi.org/10.1093/sysbio/syq010>.
39. Matsumura Y, Pitout JDD, Peirano G, DeVinney R, Noguchi T, Yamamoto M, Gomi R, Matsuda T, Nakano S, Nagao M, Tanaka M, Ichiyama S. 2017. Rapid identification of different *Escherichia coli* sequence type 131 clades. *Antimicrob Agents Chemother* 61:e00179-17. <https://doi.org/10.1128/AAC.00179-17>.
40. Hojabri Z, Darabi N, Arab M, Saffari F, Pajand O. 2019. Clonal diversity, virulence genes content and subclone status of *Escherichia coli* sequence type 131: comparative analysis of *E. coli* ST131 and non-ST131 isolates from Iran. *BMC Microbiol* 19:117. <https://doi.org/10.1186/s12866-019-1493-8>.
41. Hojabri Z, Mirmohammadkhani M, Darabi N, Arab M, Pajand O. 2019. Characterization of antibiotic-susceptibility patterns and virulence genes of five major sequence types of *Escherichia coli* isolates cultured from extraintestinal specimens: a 1-year surveillance study from Iran. *Infect Drug Resist* 12:893–903. <https://doi.org/10.2147/IDR.S199759>.
42. Hojabri Z, Mirmohammadkhani M, Kamali F, Ghassemi K, Taghavipour S, Pajand O. 2017. Molecular epidemiology of *Escherichia coli* sequence type 131 and its H30/H30-Rx subclones recovered from extra-intestinal infections: first report of OXA-48 producing ST131 clone from Iran. *Eur J Clin Microbiol Infect Dis* 36:1859–1866. <https://doi.org/10.1007/s10096-017-3021-9>.
43. Boll EJ, Overballe-Petersen S, Hasman H, Roer L, Ng K, Scheutz F, Hammerum AM, Dungu A, Hansen F, Johannesen TB, Johnson A, Nair DT, Lilje B, Hansen DS, Krogfelt KA, Johnson TJ, Price LB, Johnson JR, Struve C, Olesen B, Stegger M. 2020. Emergence of enteroaggregative *Escherichia coli* within the ST131 lineage as a cause of extraintestinal infections. *mBio* 11:e00353-20. <https://doi.org/10.1128/mBio.00353-20>.