Comparative Genome Analysis Reveals Human Pathogenic Potential of ESBL-*Escherichia coli* Isolated from Swine Microbiomes

Luria Leslie Founou[1,2,3,4*], Raspail Carrel Founou[1,4,5], Mushal Allam[6], Arshad Ismail[6], Sabiha Yusuf Essack[1,4]

[1]Antimicrobial Research Unit, University of KwaZulu-Natal, Durban, South Africa
[2]Department of Food Safety and Environmental Microbiology, Centre of Expertise and Biological Diagnostic of Cameroon (CEDBCAM), Yaounde, Cameroon
[3]Bioinformatics and Applied Machine Learning Research Unit, EDEN Foundation, Yaoundé, Cameroon
[4]AMR Insights Ambassador Network
[5]Department of Microbiology, Haematology and Immunology, Faculty of Medicine and Pharmaceutical Sciences, University of Dschang, Dschang, Cameroon
[6]Sequencing Core Facility, National Health Laboratory Service, Johannesburg, South Africa

***Corresponding author:**
Dr Luria Leslie Founou
Centre of Expertise and Biological Diagnostic of Cameroon
Yaounde, 8242
Cameroon
Email: luriafounou@gmail.com

Running title: Genome analysis of ESBL-*E. coli* in Swine

Words count: 5276

Number of Figures: 4

Number of Tables: 6

Keywords: Antibiotic resistance, ESBL-*E. coli*, Whole genome sequencing, Population Structure, Genetic Diversity, Microbiome

Abstract

Background: Extended-spectrum β-lactamase-producing *E. coli* (ESBL-Ec) harbouring virulence genes in the microbiome of food animals could likely threaten human health but is poorly understood in Cameroon and South Africa. Here, we assessed the resistome, virulome and mobilome of ESBL-Ec isolated from swine microbiomes in these countries, using whole genome sequencing (WGS).

Materials/methods: Eleven clonally-related phenotypic ESBL-Ec isolates were subjected to WGS. The isolates were *de novo* assembled using the CLC Genomics Workbench and SPAdes while RAST and PROKKA were used for annotation of the assembled contigs. Prediction of antibiotic resistance genes, virulence factors and plasmids was performed using ResFinder, VirulenceFinder and PlasmidFinder, respectively.

Results: Diverse STs were detected with sequence types ST2144 and ST88 predominating and $bla_{CTX\text{-}M\text{-}15}$ (55%) as principal ESBL genes. Although the isolates belonged mainly to commensal phylogroups A/B1 (45/28.3%) and C (18.18%), all harboured at least three extraintestinal pathogenic *E. coli* (ExPEC) VFs with one isolate harbouring up to 18 ExPEC VFs.

Conclusion: The resistance and pathogenic potential of ESBL-Ec colonizing the gut microbiota of swine in both countries demonstrate the urgent need to implement effective strategies to contain the dissemination of virulent ESBL-Ec through the food chain in Cameroon and South Africa.

Introduction

Antibiotic resistance (ABR) is a global public health issue that has severe multi-dimensional repercussions not only in humans, where decades of improvements in healthcare outcomes are threatened, but also in the food production industry. The causal relationship between the extensive usage of antibiotics in food animal production and increasing antibiotic resistance in bacteria affecting humans and animals is widely acknowledged. Antibiotics are used for a variety of purposes, including therapeutic and non-therapeutic uses as metaphylactics, prophylactics, and growth-promoters (FAO, 2015, 2016; O'Neill, 2016). Hence, the emergence and spread of ABR across the farm-to-plate continuum puts occupationally-exposed workers (viz. farmers, agricultural practitioners, abattoir workers, food handlers, etc.), their close contacts and consumers at the end of the food chain at risk of contamination or infection by antibiotic resistant bacteria (ARB) and/or antibiotic resistance genes (ARGs). ABR prevention and containment measures should focus not only on humans, but also on animals and their associated environments (Zhang et al., 2017).

Escherichia coli is a recognized commensal bacterium of the gastrointestinal tract of humans and animals. The genomic plasticity of *E. coli* strains allows their adaptation to different environments, hence their wide implication in intestinal and extraintestinal infections in both humans and animals worldwide (Sarowska et al., 2019). *E. coli* displays a clonal population structure delineating four main phylogenetic groups (A, B1, B2, and D) with pathogenic strains belonging to group B2 and D while commensal strains belong to group A and B1.

E. coli has been suggested as putative reservoir for extended-spectrum β-lactamase (ESBL) resistance and it has been demonstrated that substantial resistance emerges in commensal bacteria especially those present in the gastrointestinal tract where horizontal gene transfer

prevails and does occur within and between species and genera (Founou et al, 2016; 2019). ESBL production in *E. coli* is associated with different resistance genes but its most frequently caused by the production of ESBLs encoded by the bla_{TEM}, bla_{SHV} and bla_{CTX-M} families, with the latter being the predominant type (Perovic et al., 2014). ESBL-producing *E. coli* have been detected across the animal, human, and environmental interface worldwide, with the emergence of specific clones able to acquire ARGs and virulence genes via mobile genetic elements (MGEs) such as plasmids, transposons, gene cassettes, and other integrative genetic elements (Founou et al., 2016).

Despite the evidence of increasing prevalence of ESBL-producing *E. coli* in food animals, there is still limited information regarding the genetic structure, diversity and relationship of ESBL-*E. coli* isolates from food animals, especially in the pig industry in Sub-Saharan African countries such as Cameroon and South Africa. The objectives of this study were thus to use whole genome sequencing (WGS) and bioinformatics tools to investigate the current population structure, pathogenicity, genetic diversity and resistomes, virulomes and mobilomes, of ESBL-producing *E. coli* isolates from swine microbiomes in Cameroon and South Africa in order to ascertain their pathogenic potential in human health.

Materials and Methods

1. Ethical Considerations

Ethical approvals were obtained from the National Ethics Committee for Research in Human Health of Cameroon (**Ref. 2016/01/684/CE/CNERSH/SP**) as well as from the Biomedical Research Ethics Committee (**Ref. BE365/15**) and Animal Research Ethics Committee (**Ref. AREC/091/015D**) of the University of KwaZulu-Natal. Approvals were additionally obtained from the Cameroonian Ministry of Livestock, Fisheries and Animal Industries (**Ref.**

061/L/MINEPIA/SG/DREPIA/CE) and Ministry of Scientific Research and Innovation (**Ref. 015/MINRESI/B00/C00/C10/C14**). This study was further placed on record with the South African National Department of Agriculture, Forestry and Fisheries [**Ref.** 12/11/1/5 (878)].

2. Study Design and Bacterial Isolates

The study sample consisted of eleven putative ESBL-positive *E. coli* isolates that were collected between March and October 2016 as part of a larger study where ESBL-producing *Enterobacterales*, were collected from three abattoirs in Cameroon and two in South Africa (n=2). These isolates originating from nasal (n=6) and rectal swabs (n=5) from healthy pigs processed at abattoirs, were identified as putative, closely related ESBL producers via VITEK 2 system and enterobacterial-repetitive-polymerase chain reaction (ERIC-PCR) analysis, respectively (Founou et al., 2018).

3. Identification, ESBL Screening and Antimicrobial Susceptibility Testing

All samples were cultured on MacConkey agar supplemented with 2 mg/L cefotaxime and incubated for 18-24 h at 37^0C in normal atmosphere (Founou et al., 2019). All putative ESBL-producers, were phenotypically characterized to the genus level using Gram staining and biochemical tests (catalase and oxidase tests). The isolates were thereafter phenotypically confirmed using the VITEK 2 system.

The VITEK 2 system was further used for ESBL screening along with the double disk synergy test as previously described (Founou et al., 2019). A series of 18 antibiotics encompassed in the Vitek® 2 Gram Negative Susceptibility card (AST-N255) were tested using Vitek® 2 System and (BioMérieux, Marcy l'Etoile, France). Breakpoints of the CLSI guidelines (CLSI, 2016) were used except for the colistin, amoxicillin + clavulanic acid, piperacillin/tazobactam, amikacin for

which EUCAST breakpoints (EUCAST, 2016) were considered with *E. coli* ATCC 25922 and *K. pneumoniae* ATCC700603 being used as controls.

4. Whole genome sequencing and data analysis

4.1.Purification, Sequencing and Pre-Processing of Genomic Data

GenElute® bacterial genomic DNA kit (Sigma-Aldrich, St. Louis, MO, USA) was used for genomic DNA (gDNA) extraction with the concentration and purity assessed using agarose gel electrophoresis, NanoDrop 8000 spectrophotometer (Thermo Scientific, Waltham, MA, USA), and fluorometric analysis Qubit® (Thermo Scientific, Waltham, MA. USA). Libraries were constructed using the Nextera XT DNA Library Preparation kit (Illumina Inc., San Diego, CA, USA) and subjected to paired-end (2×300 bp) sequencing on an Illumina MiSeq (Illumina Inc., San Diego, CA, USA) machine with 100× coverage. The generated paired-end reads were merged, checked for quality, trimmed, and *de novo* assembled into contigs with the Qiagen CLC Genomics Workbench version 10.1 (CLC, Bio-QIAGEN, Aarhus, Denmark) and SPAdes version 3.11 (Bankevich et al., 2012) to overrule any inherent shortfalls from both assemblers.

4.2.WGS-based molecular typing

WGS data was used to predict *in silico* multi-locus sequence type (MLST) based on the Achtman scheme which considers allelic variation amongst seven housekeeping genes (*adk*, *fum*C, *gyr*B, *icd*, *mdh*, *pur*A and *rec*A) to assign STs (Larsen et al., 2012). In addition to generating an *E. coli* MLST assignment for each isolate, core-genome MLST (cgMLST) was assigned based on a scheme from EnteroBase server (http://enterobase.warwick.ac.uk/species/ecoli) that uses 2 513 loci (Alikhan et al., 2018). EnteroBase was further used for *in silico* phylotype predictions following the Clermont scheme (Clermont et al., 2013) as well as for *fimH* allelic designations

(Alikhan et al., 2018). Ribosomal MLST, hierarchical cgMLST clustering, wgMLST were further performed using core genome data in EnteroBase.

4.3.*In Silico* Resistome and Virulome Profiling

ARGs of the *E. coli* genomes were annotated and identified with ResFinder (Zankari et al., 2012) through the bacterial analysis online platform of GoSeqIt tool. The Comprehensive Antibiotic Resistance Database (CARD) platform was concomitantly used for prediction of ARGs and detection of chromosomal mutation (SNPs) in quinolones ARGs *gyr*A, *gyr*B, *par*C and *par*E. The selected threshold and the minimum percentage of the gene length detected were set to 90% identity for a positive match between the reference database and a target genome. VirulenceFinder (Joensen et al., 2014) available from the GoSeqIt tools server along with the comparative pathogenomics platform VFanalyzer from Virulence Factor Database (VFDB) (Liu et al., 2019) were similarly used to predict and annotate virulence factors (VFs), respectively, with a threshold of 90% identity. ExPEC virulence genes including ferric aerobactin receptor (*iutA*), increased serum survival (*iss*); heat-resistant agglutinin (*hra*), temperature sensitive haemagglutinin (*tsh*); P fimbrial adhesin (*papC*); colicin V (*cvaC*) and capsular polysialic acid virulence factor group 2 (*kpsII*) and invasive factor of brain endothelial cells locus A (*ibeA*) of *E. coli* strains responsible for neonatal meningitis in humans were investigated *in silico*. Moreover, the pathogenicity prediction web-server PathogenFinder (Cosentino et al., 2013) was used to predict bacteria pathogenic potential towards human hosts.

4.4.Detection of mobile genetic elements

The RAST SEED viewer (Overbeek et al., 2014) and Artemis Comparison Tool (ACT) were used to identify the presence of transposases and integrons flanking resistance and virulence genes. MGEFInder of the Center for Genomic and Epidemiology was used for the *in-silico*

detection of insertion sequences (IS), conjugative genetic elements and transposons allowing investigation of synteny of mobile genetic elements with VFs and antibiotic resistance genes (Durrant et al., 2020). PHAge Search Tool Enhanced Release (PHASTER) server was used for the identification, annotation and visualization of prophage sequences (Zhou et al., 2011). The profile of bacterial plasmid replicons and plasmid incompatibility groups was assessed through PlasmidFinder 2.1 (https://cge.cbs.dtu.dk/services/PlasmidFinder/) and pMLST 2.0 (https://cge.cbs.dtu.dk/services/pMLST/) (Carattoli et al., 2014). Putative CRISP system and Cas cluster were assessed through CRISPRCasFinder (https://crisprcas.i2bc.paris-saclay.fr/CrisprCasFinder/Index).

4.5.Genome visualization and gene annotation

The *de novo* assembled raw reads were annotated using the Rapid Prokaryotic Genome (PROKKA) version 1.12 beta available from EnteroBase, the NCBI Prokaryotic Genome Annotation Pipeline (PGAP) and RAST 2.0 server (http://rast.nmpdr.org)(Aziz et al., 2008) which identified encoding proteins, rRNA and tRNA, assigned functions to the genes and predicted subsystem represented in the genome. The size, GC content, average coverage, length, N50, L50, RNAs and protein coding sequences were obtained for each isolate. The annotated *in silico* predicted proteins and regions were visualized via the JSBrowser of EnteroBase server and RAST. The genomes of the isolates were visualized using the CG Viewer Server (Grant et al., 2012). In addition, the contigs were mapped against the complete genome of *E. coli Ecol_AZ155* (NZ_CP019005.1) for visualization of the genomic organization.

4.6.Comparative phylogenomic analyses

The whole genome phylogenetic relationship was assessed within the study isolates and with a collection of international *E. coli* genomes (n = 118) available at the Enterobase *E. coli* genomes

repository as of 12 November 2020. The international isolates were closely related *E. coli* strains of similar STs isolated from various sources (humans, livestock and environment).

The *E. coli* isolate YA00194039 (ERS4920643) was used as reference genome with all assembled contigs being aligned against it to determine SNP locations. The phylogeny of the *E. coli* isolates was characterised using the whole genome MLST (wgMLST), core genome MLST (cgMLST) and accessory genome MLST. Phylogenetic relationships among study isolates and between study and international isolates were assessed based on nucleotide alignments of all the genes in the entire genome (wgMLST) and core genome content (core genes that are present in most genomes with ≥95% of nucleotide identity; cgMLST). Moreover, the accessory gene including ARGs, plasmid replicons and phages content was analysed using the Enterobase server which scans the genome against the core ResFinder and PlasmidFinder databases based on a percentage identity of ≥ 90% and coverage of ≥ 70% in order to generate a customized phylogenetic tree to infer the evolutionary relationship within the study isolates and between the study and international isolates. GrapeTree minimum spanning and phylogenetic trees were further built to describe the relatedness among the study isolates and between the study and international isolates. The generated phylogenomic trees were downloaded, and subsequently visualized and edited using MicroReact (www.microreact.org).

Results

1. Baseline characteristics and phenotypic analyses

All the *E. coli* isolates were ESBL producers and high level of resistance to amoxicillin, cefuroxime, cefuroxime axetil, as well as to third (cefotaxime, ceftazidime) and fourth generation (cefepime) cephalosporins were observed. All isolates were resistant to trimethoprim-sulfamethoxazole and susceptible to cefoxitin, ertapenem, meropenem, imipenem and

tigecycline. Altogether, three resistance patterns were identified with the pattern AMP.AMC.TZP.CXM.CXM-A.CTX.CAZ.FEP.TMP/SXT (82%) being the most prevalent one. Two isolates displayed multidrug resistance (MDR; resistance to three or more antibiotic families) with one isolate, PN256E8, being resistant to colistin with a minimum inhibitory concentration (MIC) of 8 mg/L. All but two isolates were susceptible to nitrofurantoin. Relevant population data, specimen source, phenotypic and genotypic characteristics for these isolates are summarized in Supplementary Table 1.

2. Genomic features

Table 1 and Table S1 depict all the genomic characteristics including length, GC content, N50, coverage, coding sequences, RNAs, rMLST, phylotype, serotype, CRISPR arrays, etc. of the isolates. The genome size of the isolates ranged from 4.5 Mb to 5.3 Mb with a GC content of 50.5 to 50.9 and coverage of 111 to 188.

3. Antimicrobial resistance phenotypes and genotypes

Whole genome-based resistomes corroborated the observed phenotypic profiles. All isolates evidenced relatively similar combinations of resistance genes encoding target modification, antibiotic inactivation, antibiotic efflux pumps and regulators. All *E. coli* isolates harboured $bla_{CTX\text{-}M}$ with $bla_{CTX\text{-}M\text{-}15}$ (n=6; 54.54%) being the most frequently identified. The $bla_{CTX\text{-}M\text{-}14}$ (PR209E1, PR246B1C) was detected in two isolates as was the $bla_{CTX\text{-}M\text{-}1}$ (PN256E1, PN256E2). Three isolates simultaneously harboured the $bla_{CTX\text{-}M\text{-}15}$ and $bla_{TEM\text{-}1B}$ whilst one isolate, PN256E8 harboured concomitantly the $bla_{CTX\text{-}M\text{-}15}$, $bla_{TEM\text{-}1B}$, $bla_{TEM\text{-}141}$ and $bla_{TEM\text{-}206}$ (Table 2). Although only two isolates (PR010E3, PN256E8) displayed phenotypic resistance to aminoglycosides, all except two isolates (PR256E1, PN256E2) harboured aminoglycoside resistance genes including str, *aac*, *aad* and *aph*. Specifically, *aph* was identified in six isolates:

246 *aph(3'')-Ib* (6/11, 55%) and *aph(6)-1d* (6/11, 55%) genes (Table 2). Eight (73%) isolates
247 harboured different *aad* genes; including *aadA5 (3/11; 27%), aadA1 (3/11; 27%)* and
248 *aadA2(2/11; 18.18)* (Tables 1). Four (36%) isolates harboured *aac* genes including *aac(3)-IIa*
249 (1/11; 9%), and *aac(6)-Ib* (3/11; 27%).
250 Several types of plasmid-mediated quinolone resistance (PMQR) genes were identified in the
251 isolates (Table 2). Specifically, the *QnrS1* gene was present in four (36%) isolates, whilst the
252 *aac(6')Ib-cr* and OqxAB genes were identified in three isolates (27%) each. Mutations in the
253 *gyrA* quinolone resistance-determining region (*QRDR*) genes was observed in two isolates
254 (PR010E3, PN091E1II). *gyrA* had three mutations with two (S**83L**, D**87**N) occurring within
255 PR010E3 and one in PN091E1II (S**83**A) (Table 2). All isolates, except for PN256E8 and
256 PR010E3, for which PMQR gene was identified in both and additionally QRDR in the latter,
257 were susceptible to ciprofloxacin.
258 All isolates displayed concomitant resistance to trimethoprim and sulfamethoxazole with all
259 harbouring *sul* genes. Specifically, *sul2* gene was identified in nine (82%) isolates alone and in
260 combination with *sul1* gene in one (PR010E3). The rarely evidenced *sul3* gene, was identified in
261 two isolates (PR246B1C, PR209E1). Similarly, the *dfr* gene was identified in 7/11 (64%)
262 isolates, specifically *dfrA17* (n=3), *dfrA14* (n=3) and *dfrA1* (n=1). Diverse permutations of *dfr*
263 and *sul* genes occurred in the isolates. For instance, *sul2* and *dfrA14* were identified in 3/11
264 (27%) while *sul2* and *dfrA17* were detected in two isolates. One isolate, PR010E3 harboured
265 *sul2*, *sul1* and *dfrA17* concomitantly. In the two isolates harbouring the *sul3* gene, the
266 fosfomycin resistance gene *fosA3* was also detected along with the chloramphenicol resistance
267 gene *cmlA1*. Likewise, in the two isolates harbouring *dfrA17*, the florfenicol resistance gene floR
268 was also detected (Table 2). One isolate (9%) harboured the mcr-1 gene in a full-length copy of a

colistin resistance gene showing 100% nucleotide similarity to the reference database sequence (Table 2).

4. Whole-genome virulome profiling and pathogenicity

The virulomes of all *E. coli* displayed high level of pathogenicity (Table 3). The ESBL-*E. coli* isolates showed a 93.6% mean probability (P score) of being human pathogens. The pathogenic species with the highest linkage (100% identity) were the *E. coli* APEC O1 (Accession numbers: DQ517526, DQ381420), *E. coli* UMN026 (Accession number: CU928163) and *E. coli* UTI89 (Accession number: CP000243) which are all extraintestinal pathogenic strains in animals (poultry) and humans belonging to the pathogenic phylogroup B2.

Altogether, over 470 putative virulence factors (VFs) were detected in all isolates. The VFs belonged to major functional categories including: adhesins, toxins, protectins and invasins, iron uptake/siderophores, anti-phagocytosis, secretion systems and autotransporters (Table 3). The isolate PR85E1 harboured the highest number (133) of VFs, followed by the isolates PN256E2 and PR010E3 with 76 and 63 VFs, respectively. All except PR246B1C, PN256E8 and PN027E6II isolates had more than 20 VFs. Analysis of the virulence genotype *fimH* showed that it was present in 64% of the isolates. Among putative VFs, autotransporter adhesin *ehaB*, invasin of brain endothelial cells locus B (*ibeB)* and invasin of brain endothelial cells locus locus C (*ibeC)*, belonging to autotransporter protein and invasins, were the most prevalent (>73%, 8/11) VFs across the isolates (Tables 3-4). The episomal *iss* gene was detected in 55% of isolates whilst *papC,* periplasmic iron-binding protein *sitA (sitA),* siderophore yersiniabactin receptor *(fyuA) genes* were identified in 27% (3/11) isolates. Similarly, the prevalence of protectin genes including the complement resistance protein *traT* was the same as that of iron acquisition genes *iroN and iutA* (Tables 3-4). Interestingly, the avian hemolysin gene F (*hlyF)* that enhance

292 production of outer membrane vesicles (OMVs) and lead to autophagy of eukaryotic cells was
293 detected in one isolate while the hemolysin E (*hlyE*) a pore-forming toxin was observed in 34%
294 (4/11) isolates.
295 Our findings showed that rectal *E. coli* isolates harboured more VFs than nasal isolates with
296 prevalence of VFs ranging from 6 to 133 against 4 to 76 in nasal isolates. Putative VFs for
297 invasion, such as the outer membrane protein T (OmpT*)* and the *traT* genes were most prevalent
298 in rectal than nasal isolates. However, the polysialic acid transport protein group 3 (*KpsMIII)*
299 gene encoding for group 3 capsule, was detected only in nasal isolates as were the unique *vat* and
300 *astA*. Specifically, all *E. coli* isolates harboured at least one ExPEC VF from each of the major
301 functional categories and up to 18 ExPEC VFs (Table 3).

302 **5. Phylogenetic groups and multilocus sequence typing, serotyping and phylotyping**

303 Based on *in silico* MLST results, four *E. coli* isolates were assigned to the pandemic ST88 (n=2)
304 and ST2144 (n=2) clones, while the remaining isolates were assigned to six single-locus variants,
305 namely, ST10, ST69, ST226, ST944, ST4450 and ST44. Interestingly, the two *E. coli* ST88
306 strains, PN256E2 and PR256E1, were detected in nasal and rectal pools of the same pig while
307 the *E. coli* ST2144 were both isolated from two rectal pools of swine originating from the same
308 farm and processed within the same abattoir (SH004).
309 The majority of the isolates were assigned to commensal phylogroups A (45%), B1 (28%) and C
310 (18%) but one belonged to the virulence phylogroup D (9%). The serotype O-:H49 (18.18%) and
311 O-:H18 (18.18%) were the principal serotypes detected while the *fimH*1250 (18.18%) and
312 *fimH*87 (18.18%) were the predominant *fimH* gene observed.

313 **6. Mobile genetic elements**

WGS analysis identified 15 different plasmid replicons in all the isolates, which further all harboured multiple plasmid replicons concomitantly. Ten types of incompatibility (Inc) plasmid replicons were identified with different frequencies including IncY, IncFIA, IncFIB (AP001918), IncFIC(FII), IncFII, IncN, IncHI2, IncHI2A, IncI1, IncI2 and IncX (Tables 2 and 5). The majority of isolates (9/11; 82%) harboured the IncF (FII, FIB, FIC, FIA) and IncY (4/11; 36%). Four isolates harbouring the IncF incompatibility group also harboured the IncH (n=2) and IncI (n=2) groups.

In silico plasmid MLST-analyses assigned the IncF plasmid incompatibility group to STs K-:A-:B1 and K89:A-:B57 while IncH and IncI plasmids were assigned to ST3. Additionally, nine (82%) isolates harboured an array of insertion sequences (IS) with IS26, IS421, Isec1 being the most frequent with a 55% prevalence. An array of three IS (IS26, ISVsa3, ISEc9) were harboured on the plasmid IncI that also encoded the sulphonamide resistance gene (*sul2*) and virulence factor (*cib*) in the two *E. coli* ST88 (PR256E1 and PN256E2). Similarly, three isolates harboured transposons (Tn) including Tn6082 (18%) and Tn7 (9%), with the trimethoprim resistance gene *dfrA1* being encoded in transposon Tn7 (Table 5).

7. Phylogenetic analysis

The contigs of the ESBL-*E. coli* harbouring the most VFs were mapped against the complete genome of *E. coli* Ecol_AZ155 (NZ_CP019005.1) for visualization of the genomic organisation (Figure 1). Results of comparative genomic analyses revealed specific similarities and dissimilarities, i.e., isolates had similar and dissimilar arrangements of genomic regions towards representative and reference genomes (Figure 1).

Whole genome phylogenetic analysis grouped the study *E. coli* isolates (n = 11) into two major clusters (Figure 2). The first one grouped six isolates including two *E. coli* ST2144 (PR209E1,

PR246B1C), two ST88 strains (PR256E1, PN256E2), one ST940 (PN091E1II) and one ST4450 (PR085E3). The two *E. coli* ST2144 isolates identified from the same abattoir (SH004) in South Africa had 100% identity and shared a close common ancestor with *E. coli* ST940 (PN091E3) and *E. coli* ST4450 (PR085E3) which both originated from one Cameroonian abattoir (SH002). Similarly, the two ST88 isolates (PR256E1 and PN256E2) displayed 100% identity and were found to share common ancestor with the *E. coli* ST2144, ST940 and ST4450. The second clade included four isolates belonging to various STs (i.e., PN256E8: ST9440; PR010E3: ST44; PN017E1: ST10; PN027E1II: ST226) of which *E. coli* ST10 (PN017E1) and ST44 (PR010E3) isolated in the same abattoir Cameroon were closely related and shared common ancestor with *E. coli* ST9440 originating from South Africa.

The phylogenomic analyses of the isolates with international strains revealed that they were more closely related to strains from Africa including Kenya, Ghana, Egypt and Morocco than to any other country. None were phylogenetically related to any strain from United States and United Kingdom. The genomes from livestock and humans clustered together at some level. Specifically, PR246B1C and PR209E1 (ST2144), described above to have the same virulome, resistome and mobilome, as well as PN091E1II (ST940), were in the same cluster and shared common ancestors with strains isolated from humans and livestock in Kenya, Uganda, South Africa, Mozambique and Ghana (Figure 3). Similarly, hierarchical clustering analyses provided evidence of genomic relationship between strains originating from livestock and humans with the sole *E. coli* ST10 strain (PN017E2) sharing common ancestor with *E. coli* ST10 isolated from human in Spain and livestock in Luxembourg. Intriguingly, the ST9440 isolate (PN256E8) was closest to a ST10 isolate (Ec416) from livestock in Vietnam. Phylogenetic tree based on

accessory genome within the study isolates and between the international isolates revealed similar findings (Figure 4).

Discussion

ESBL-*E. coli* is responsible of severe infections worldwide. However, genotypic and pathogenic characteristics of isolates from food-producing animals are not well documented in Cameroon and South Africa. In this study, the population structure of ESBL-*E. coli* isolated from swine microbiomes collected at Cameroonian and South African abattoirs were investigated using WGS. The swine microbiomes from these settings harboured a diverse population of *E. coli* strains that encode an extensive repertoire of resistance genes and plethora of virulence factors, and harboured mainly in IncF, IncH and IncY plasmid replicons.

The antimicrobial resistance data clearly indicates high resistance level of ESBL-*E. coli* isolates to ampicillin, cefuroxime, cefotaxime, ceftazidime and trimethoprim-sulfamethoxazole. Antibiotic use should be limited in agricultural practices and food production systems in order to curtail the emergence of resistance. Each genome of ESBL-*E. coli* isolated from both countries harboured chromosomal and plasmid-mediated ARGs. The ESBL phenotype of the isolates has been confirmed by the presence of ESBL genes ($bla_{CTX\text{-}M}$, $bla_{OXA\text{-}1}$ and $bla_{TEM\text{-}1B}$) supporting other contemporary studies which showed that blaCTX-M-15 is the most prevalent ESBL variants among *E. coli,* especially in the highly virulent clone ST131 (Mathers et al., 2015; Rafaï et al., 2015). In fact, Rafai et al. (2015) detected 63.7% of ESBL producers in surgical site infections in Central African Republic. The authors showed that blaCTX-M-15 was present in all isolates along with aac(6')-Ib-cr. Similar finding were also evidenced by Mbelle et al. (2019) who reported a 70% of blaCTX-M-15 among hospitalized patients in South Africa.

A major observation was the phenotypic multi-drug resistance (MDR) of the isolates PR010E3 and PN256E8 although all isolates, except PR256E1 and PN256E2, harboured concomitantly resistance genes encoding for resistance to aminoglycosides, tetracyclines and fluoroquinolones. This finding is in contrast with data readily available from the literature which suggest that machine or deep learning can be used to predict adequately AMR and MIC of antimicrobial based on genome sequence data (Hendriksen et al., 2019). While transcription analyses could not be undertaken to assess the expression of these genes, these observations led us to posit that the PMQR and QRDR genes as well as aminoglycoside resistance genes present in these isolates might not have been expressed or be silent. Similar discrepancies regarding the phenomes and genomes of isolates harbouring resistance genes but not expressing associated phenotypic resistance were reported elsewhere (Mbelle et al., 2019) and can be further observed in Tables 2 and S2. Our finding further reveals that application of machine or deep learning as well as comparison between phenome and genome are still needed at a large-scale from various environments and sources.

All isolates were seen as human pathogens with over 93% of pathogenicity score and *E. coli* APEC-O1-ColBM (DQ381420), *E. coli* UTI89 (CP000243) and *E. coli* K12 DH10B (CP000948) being the closest related strains. There is increasing evidence that food producing animals and food products or animal origin may contribute to the spread of ExPEC in the community (Hammad et al., 2019). In our study, all ESBL-*E. coli* harboured at least three VFs associated with ExPEC such as *iss*, *iutA*, *traT*, *ompT*, *hlyA*, *iroN*, *papC* and *fimH*. Of great concern is that the isolate PR085EE3 carried over to 100 VFs with the majority having been identified in clinical ExPEC. This suggests that commensal bacteria prevailing in the microbiome of food animals are not only reservoir of resistance genes, but more so, of virulence factors which might

be transmitted via HGT and spread to humans through the food chain. It reemphasizes the need to ensure adequate food safety measures throughout the farm-to-plate continuum along with effective infection prevention and control measures in hospitals.

Biofilm formation, essential in colonization and resistance to antibiotics, is an important virulence factor in bacteria. Although biofilm-associated genes have been found to be associated to phylogroup B2 and D strains, biofilm-associated genes such as *fim* and *pap* clusters were detected in our isolates despite being mainly of phylogroup A and B1. We posit that this might again be due to HGT of genetic factors which might have allowed our isolates, though commensals, to acquire these virulence genes horizontally and become putatively virulent in case of extra-intestinal infections. Moreover, the detection of the heat stable enterotoxin 1 (*astA*) gene, encoding for the enteroaggregative *E. coli* heat-stable toxin 1 (EAST1), in the sole ST9440 strain (PN256E8), as well as the avian hemolysin (*hlyF*) in PR256E1 (ST88) gives credence to the fact that commensal *E. coli* prevailing in the gut microbiome have a propensity to acquire various virulence genes which might therefore evolve as progenitor lineages from which heteropathogenic *E. coli* including uropathogenic *E. coli* (UPEC), neonatal meningitis-associated *E. coli* (NMEC) and enteroaggregative *E. coli* (EAEC) strains will emerge.

The majority of commensal *E. coli* strains belong to the phylogenetic groups A and B1, whereas the most common virulent ExPEC are associated with group B2 and D. Our results revealed that the majority of ESBL-*E. coli* isolates belonged to group A (45%) and B1 (28%). Several reports confirmed that phylogroups A and B1 are the leading phylogroups among *E. coli* isolates especially in the gut microbiome (Li et al., 2010; Stoppe et al., 2017). A study from Nigeria showed that 62% of *E. coli* isolates tended towards the commensal phylogroup B1 and A (Olowe et al., 2019). The relationship between phylogenetic groups and ABR has been established

previously (Olowe et al., 2019) and studies have shown that the group B2 strains are mainly MDR. However, our study revealed that isolates from other phylogroups such as group A and B1 could also display MDR and high level of virulence certainly as a result of HGT.

Despite the limited sample size (11 *E. coli* isolated from 8 pooled samples), our findings reflect the rich diversity existing within *E. coli* population structure with nine sequence types being detected. The ESBL- *E. coli* isolates were mainly circulating in two clonal lineages since four out of seven isolated strains belong to the ST2144 (n=2) and ST88 (n=2). In addition, the MDR-high-risk clone ST69 and the ST10 were also detected. The ST10 complex, including ST10, commonly associated with spread of CTX-M-1, CTX-M-2 and CTX-M-9 groups, is highly distributed among humans and various livestock species and has been linked with intestinal and extraintestinal infections in several African countries (Rafaï et al., 2015). *E. coli* ST10 was the main ST along with ST131 identified in surgical site infection in Central African Republic (Rafaï et al., 2015). Like other high-risk clones, *E. coli* ST69 possesses biological factors such as *usp*, *ompT*, secreted autotransporter toxin (*sat*) and *iutA* genes corresponding specifically to ST131 (Mathers et al., 2015), that increases bacterial fitness allowing these strains to out-compete other bacterial strains and become the principal part of the bacterial population in the gut (Mathers et al., 2015). The detection of mcr-1 gene in one isolate suggests that colistin resistant *Enterobacterales* are also emerging among food-producing animals in Africa, and demonstrates the urgent need of antimicrobial usage stewardship in food production systems and implementation of effective monitoring programmes to curb the spread of MDR-*E. coli*.

Comparative hierarchical clustering suggested that the majority of our strains belong to two clusters. It further confirmed that ST10 complex is common in African livestock as all our ST10 complex belong to a unique cgMLST cluster containing closely related isolates from

Cameroonian (PR010E3, PN027E1II) and South African swine (PN256E8). The comparative phylogenomic analysis further confirms that our ESBL-*E. coli* ST10 demonstrated overlap with ST10 strains isolated from livestock and human populations in Africa (Mozambique, Egypt, Morocco), Asia (Vietnam), Europe (United Kingdom, Denmark, France) and Oceanie (Australia) and displaying the same phylogroup A. Similarly, our ESBL-*E. coli* ST69 share high level of similarity with an UPEC ST69 phylogroup D that was involved in pyelonephritis in France (unpublished data). This gives credence to the hypothesis that commensal *E. coli* of the gut microbiome might be the reservoir of virulence and resistance genes that allow the emergence of hetero-pathogenic *E. coli* strains.

Mobile genetic elements (MGEs) play an essential role in the mobility of ARGs and VFs between different bacterial species. The plasmids of the Inc-family are commonly associated with MDR and VFs in *E. coli* (Rafaï et al., 2015). Ten (91%) of our ESBL-*E. coli* harboured Inc-related replicons with IncF (9/11; 81%) especially the replicon *FIB* (AP001918; 46%) being the leading, followed by IncY (4/11, 36%) and IncH (3/11; 21%) plasmids. Similar plasmid replicons associated with blaCTX-M-group were reported in humans and livestock in Africa and across the world (Patil, Chen, Lian, & Wen, 2019).

Our *E. coli* isolates harboured multiple plasmids belonging to major replicon types and encoding multiple ARGs and VFs. Given the presence of ESBL-*E. coli* in clinically healthy animals and humans, it is likely that the presence of these plasmids could contribute to the long-term persistence of resistance traits in animal and environmental microbiome. Though our study was limited by the isolates numbers and geographic area, our results sufficiently reinforce the need to closely monitor pathogenic and commensal bacteria prevailing in the food production systems on the continent (Patil et al., 2019).

Conclusion

Our study demonstrates that the population structure of ESBLs-*E. coli* in swine microbiome is highly diverse with the *bla*$_{CTX\text{-}M\text{-}15}$ gene being the leading CTX-M variant. Although the phylogenetic diversity observed precludes any suggestion for clonal dissemination, the resistance and high human pathogenic potential demonstrate the urgent need to implement effective strategies to contain the dissemination of antibiotic resistant bacteria in Cameroon and South Africa. Our study underlines the necessity of long-term and stringent monitoring and molecular studies investigating population structure of commensal and pathogenic bacteria in (food) animals, food products and associated environments, not only to preserve antibiotics for future generations, but also to gain new insights into the diversity, evolutionary history and emergence of ESBL-ExPEC. The Africa Pathogen Genomics Initiative (Africa PGI, https://africacdc.org/download/africa-pathogen-genomics-initiative-factsheet/) recently launched, and aiming at integrating pathogenomics and bioinformatics into outbreak investigations and public health surveillance is expected to advance molecular-based studies and improved infection control and prevention on the continent.

Acknowledgements

The authors would like to thank all included abattoirs, participants and other involved employees from Cameroon and South Africa for participation and all field work assistance.

Preliminary results from this study were presented in a poster at the 30th European Congress of Clinical Microbiology and Infectious Diseases (ECCMID) in Paris, France, 2020 (abstract 2297).

Funding

L.L. Founou and **R.C. Founou** are funded by the Antimicrobial Research Unit (ARU) and College of Health Sciences (CHS) of the University of KwaZulu-Natal, South Africa. The National Research Foundation funded this study through the NRF Incentive Funding for Rated Researchers (Grant No. **85595**), and the DST/NRF South African Research Chair in Antibiotic Resistance and One Health (Grant No. **98342**) awarded to **S.Y. Essack**. The South African

Medical Research Council also funded the study through the Self-Initiated Researcher (SIR) Grant awarded to **S.Y Essack**. The funders had no role in the study design, preparation of the manuscript nor the decision to submit the work for publication.

Conflict of interest

Professor Essack is a member of the Global Respiratory Infection Partnership and Global Hygiene Council both sponsored by unrestricted educational grants from Reckitt and Benckiser Ltd., UK. All other authors declare that there is no competing financial interest.

Author contributions

L.L.F co-conceptualized the study, undertook sample collection, microbiological laboratory and data analyses, prepared tables and figures, interpreted results, performed bioinformatics analysis, and drafted the manuscript. **R.C.F** undertook sample collection, microbiological laboratory analyses, contributed to bioinformatics analysis, vetting of the results and writing of the manuscript. **M.A.** contributed to and vetted bioinformatics analyses. **A.I.** performed whole genome sequencing analysis. **S.Y.E** co-conceptualized the study and undertook critical revision of the manuscript. All authors read and approve the final manuscript.

References

Alikhan, N., Zhou, Z., Sergeant, M., & Achtman, M. (2018). A genomic overview of the population structure of Salmonella. *PLoS Genet, 14*(4), e1007261. doi:https://doi.org/10.1371/journal.pgen.1007261

Aziz, R. K., Bartels, D., Best, A. A., DeJongh, M., Disz, T., Edwards, R. A., et al. (2008). The RAST Server: rapid annotations using subsystems technology. *BMC Genomics, 9*, 75. doi:10.1186/1471-2164-9-75

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., et al. (2012). SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *J Comput Biol, 19*(5), 455-477. doi:10.1089/cmb.2012.0021

Carattoli, A., Zankari, E., García-Fernández, A., Voldby, L., Lund, O., Villa, L., et al. (2014). In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. *Antimicrob Agents Chemother, 58*(7), 3895-3903.

Clermont, O., Christenson, J. K., Denamur, E., & Gordon, D. M. (2013). The Clermont Escherichia coli phylo-typing method revisited: improvement of specificity and detection of new phylo-groups. *Environ Microbiol Rep, 5*, 58-65. doi:http://dx.doi.org/10.1111/1758-2229.12019

CLSI. (2016). *Performance Standards for Antimicrobial Susceptibility Testing; Twenty-Six Supplement M100S. CLSI document M100-S24* Retrieved from WAYNE, USA:

Cosentino, S., Voldby Larsen, M., Møller Aarestrup, F., & Lund, O. (2013). PathogenFinder - Distinguishing Friend from Foe Using Bacterial Whole Genome Sequence Data. *PLoS ONE, 8*(10), e77302. doi:10.1371/journal.pone.0077302

Durrant, M. G., Li, M. M., Siranosian, B. A., Montgomery, S. B., & Bhatt, A. S. (2020). A Bioinformatic Analysis of Integrative Mobile Genetic Elements Highlights Their Role in Bacterial Adaptation. *Cell Host & Microbe, 27*(1), 140-153.e149. doi:https://doi.org/10.1016/j.chom.2019.10.022

EUCAST. (2016). *Breakpoint tables for interpretation of MICs and zone diameters* Retrieved from http://www.eucast.org

Founou, L. L., Founou, R. C., & Essack, S. Y. (2016). Antibiotic Resistance in the Food Chain: A Developing Country-Perspective. *Front Microbiol, 7*(1881). doi:10.3389/fmicb.2016.01881

Founou, L. L., Founou, R. C., Ntshobeni, N., Govinden, U., Bester, L. A., Chenia, H. Y., et al. (2019). Emergence and Spread of Extended Spectrum beta-Lactamase Producing Enterobacteriaceae (ESBL-PE) in Pigs and Exposed Workers: A Multicentre Comparative Study between Cameroon and South Africa. *Pathogens, 8*(1). doi:10.3390/pathogens8010010

Hammad, A. M., Hoffmann, M., Gonzalez-Escalona, N., Abbas, N. H., Yao, K., Koenig, S., et al. (2019). Genomic features of colistin resistant Escherichia coli ST69 strain harboring mcr-1 on IncHI2 plasmid from raw milk cheese in Egypt. *Infect Genet Evol, 73*, 126-131. doi:https://doi.org/10.1016/j.meegid.2019.04.021

Hendriksen, R. S., Bortolaia, V., Tate, H., Tyson, G. H., Aarestrup, F. M., & McDermott, P. F. (2019). Using Genomics to Track Global Antimicrobial Resistance. *Front Public Health, 7*(242). doi:10.3389/fpubh.2019.00242

Joensen KG, Scheutz F, Lund O, Hasman H, Kaas RS, Nielsen EM, & FM, A. (2014). Real-time whole-genome sequencing for routine typing, surveillance, and outbreak detection of verotoxigenic Escherichia coli. *J Clin Microbiol, 52*(5), 1501-1510.

Larsen, M., Cosentino, S., Rasmussen, S., Friis, C., Hasman, H., Marvig, R. L., et al. (2012). Multilocus sequence typing of total-genome-sequenced bacteria. *J Clin Microbiol, 50*, 1355-1361. doi:http://dx.doi.org/10.1128/JCM.06094-11

Li, B., Sun, J.-y., Han, L.-z., Huang, X.-h., Fu, Q., & Ni, Y.-x. (2010). Phylogenetic Groups and Pathogenicity Island Markers in Fecal Escherichia coli Isolates from Asymptomatic Humans in China. *Appl Environ Microbiol, 76*(19), 6698-6700. doi:10.1128/aem.00707-10

Liu, B., Zheng, D., Jin, Q., Chen, L., & Yang, J. (2019). VFDB 2019: a comparative pathogenomic platform with an interactive web interface. *Nucleic Acids Res, 47*(D1), D687-d692. doi:10.1093/nar/gky1080

Mathers, A. J., Peirano, G., & Pitout, J. D. D. (2015). The Role of Epidemic Resistance Plasmids and International High-Risk Clones in the Spread of Multidrug-Resistant Enterobacteriaceae. *Clin Microbiol Rev, 28*(3), 565. doi:10.1128/CMR.00116-14

Mbelle, N., Feldman, F., Sekyere, J., Maningi, N., Modipane, L., & Essack, S. (2019). The Resistome, Mobilome, Virulome and Phylogenomics of Multidrug-Resistant Escherichia coli Clinical Isolates from Pretoria, South Africa. *Sci Rep*. doi:10.1038/s41598-019-52859-2

Musicha, P., Feasey, N. A., Cain, A. K., Kallonen, T., Chaguza, C., Peno, C., et al. (2017). Genomic landscape of extended-spectrum β-lactamase resistance in Escherichia coli from an urban African setting. *J Antimicrob Chemother, 72*(6), 1602-1609. doi:10.1093/jac/dkx058

Namboodiri, S. S., Opintan, J. A., Lijek, R. S., Newman, M. J., & Okeke, I. N. (2011). Quinolone resistance in Escherichia coli from Accra, Ghana. *BMC Microbiology, 11*, 44-44. doi:10.1186/1471-2180-11-44

Olowe, O. A., Adefioye, O. J., Ajayeoba, T. A., Schiebel, J., Weinreich, J., Ali, A., et al. (2019). Phylogenetic grouping and biofilm formation of multidrug resistant Escherichia coli isolates from humans, animals and food products in South-West Nigeria. *Sci Afr, 6*, e00158. doi:https://doi.org/10.1016/j.sciaf.2019.e00158

Overbeek, R., Olson, R., Pusch, G. D., Olsen, G. J., Davis, J. J., Disz, T., et al. (2014). The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Res, 42*(Database issue), D206-214. doi:10.1093/nar/gkt1226

Patil, S., Chen, X., Lian, M., & Wen, F. (2019). Phenotypic and genotypic characterization of multi-drug-resistant Escherichia coli isolates harboring bla(CTX-M) group extended-spectrum β-lactamases recovered from pediatric patients in Shenzhen, southern China. *Infect Drug Resist, 12*, 1325-1332. doi:10.2147/IDR.S199861

Perovic, O., Singh-Moodley, A., Duse, A., Bamford, C., Elliott, G., Swe-Han, K. S., et al. (2014). National sentinel site surveillance for antimicrobial resistance in Klebsiella pneumoniae isolates in South Africa, 2010-2012. *S Afr Med J, 104*(8), 563-568. doi:10.7196/samj.7617

Rafaï, C., Frank, T., Manirakiza, A., Gaudeuille, A., Mbecko, J.-R., Nghario, L., et al. (2015). Dissemination of IncF-type plasmids in multiresistant CTX-M-15-producing Enterobacteriaceae isolates from surgical-site infections in Bangui, Central African Republic. *BMC Microbiology, 15*(1), 15. doi:10.1186/s12866-015-0348-1

Sarowska, J., Futoma-Koloch, B., Jama-Kmiecik, A., Frej-Madrzak, M., Ksiazczyk, M., Bugla-Ploskonska, G., & Choroszy-Krol, I. (2019). Virulence factors, prevalence and potential transmission of extraintestinal pathogenic Escherichia coli isolated from different sources: recent reports. *Gut Pathog, 11*, 10-10. doi:10.1186/s13099-019-0290-0

Stoppe, N. d. C., Silva, J. S., Carlos, C., Sato, M. I. Z., Saraiva, A. M., Ottoboni, L. M. M., & Torres, T. T. (2017). Worldwide Phylogenetic Group Patterns of Escherichia coli from Commensal Human and Wastewater Treatment Plant Isolates. *Frontiers in Microbiology, 8*, 2512-2512. doi:10.3389/fmicb.2017.02512

Yang, F., Zhang, S., Shang, X., Wang, L., Li, H., & Wang, X. (2018). Characteristics of quinolone-resistant Escherichia coli isolated from bovine mastitis in China. *Journal of Dairy Science, 101*(7), 6244-6252. doi:https://doi.org/10.3168/jds.2017-14156

Zankari, E., Hasman, H., Cosentino, S., Vestergaard, M., Rasmussen, S., Lund, O., et al. (2012). Identification of acquired antimicrobial resistance genes. *J Antimicrob Chemother, 67*(11), 2640-2644.

Zhou, Y., Liang, Y., Lynch, K. H., Dennis, J. J., & Wishart, D. S. (2011). PHAST: A Fast Phage Search Tool. *Nucleic Acids Research, 39*(suppl_2), W347-W352. doi:10.1093/nar/gkr485

Figure Legends

Figure 1. Circular genome representation of selected ESBL-*E. coli* aligned with reference genome and closely related strains.

Figure 2. Comparative genome analysis based on the core genome MLST of *E. coli* isolates.

Figure 3. Comparative genome analysis based on the core genome MLST compared with international *E. coli* isolates.

Figure 4. Comparative genome analysis based on the accessory genome of ESBL-*E. coli* isolates

Table 1. Genotypic characteristics of ESBL-producing *E. coli* isolates (Bioproject PRJNA412434)

Isolate	Accession Number	Country	Sample type	Abattoir	MLST*	Clonal Complex	FimH sub-typing	Phylogroup	Serotype
PN017E2II	VMKK00000000	Cameroon	Nasal swab	SH001	10	ST10 Cplx	FimH215	A	O9:H:9
PR010E3I	VKOQ00000000	Cameroon	Rectal swab	SH001	44	ST10 Cplx	FimH54	A	O89:H4
PN027E6IIB	VKOV00000000	Cameroon	Nasal swab	SH001	69	ST69 Cplx	FimH27	D	O-:H18
PR256E1	VKOS00000000	South Africa	Rectal swab	SH005	88	ST23 Cplx	FimH1250	C	O: Uncertain H9
PN256E2	VKOT00000000	South Africa	Nasal swab	SH005	88	ST23 Cplx	FimH1250	C	O-:H9
PN027E1II	VKOW00000000	Cameroon	Nasal swab	SH001	226	ST226 Cplx	FimH43	A	O-:H19
PN091E1II	VKOU00000000	Cameroon	Nasal swab	SH002	940	ST448 Cplx	Unknown	B1	O-:H33
PN256E8	QJRZ00000000	South Africa	Nasal swab	SH005	9440	ST10 Cplx	FimH23	A	O-:H52
PR209E1	VKOO00000000	South Africa	Rectal swab	SH004	2144	-	FimH87	B1	O-:H49
PR246B1C	WHRW00000000	South Africa	Rectal swab	SH004	2144	-	FimH87	B1	O-:H49
PR085E3	VKOP000000000	Cameroon	Rectal swab	SH002	4450	-	FimH566	A	O-:H18

Table 2. Overview of resistome and mobilome in ESBL-producing *E. coli* isolates

Isolate	Country	Sample type	Abattoir	MLST*	β-Lactamase resistance genes			Fluoroquinolone resistance genes		Other resistance genes	Plasmids	pMLST*
					BlaCTX-M	*BlaTEM*	*BlaOXA*	QRDR	PMQR			
PN017E2II	Cameroon	Nasal swab	SH001	10	*CTX-M-15*	*TEM-1B*	-	-	*QnrS1*	*aph(6)-Id, aph(3'')-Ib, tet(A), mph(A), sul2, dfrA14,*	IncY, Col(MG828), Col440I, rep21	-
PR010E3I	Cameroon	Rectal swab	SH001	44	*CTX-M-15*	-	*OXA-1*	*gyrA (p.S83L) gyrA (p.D87N)*	*aac(6')-Ib-cr*	*aac(3)-IIa, aph(3'')-Ib, aadA5, aph(6)-Id, tet(B), tet(A), sul1, sul2, dfrA17, floR, catB3*	IncFIA, Col440I, IncFII, IncFIB, Col(MG828), rep21	IncF [F36:A20:B1]
PN027E6IIB	Cameroon	Nasal swab	SH001	69	*CTX-M-15*	*TEM-1B*	-	-	*QnrS1*	*strA, strB, sul2, tet(A), dfrA14*	IncY, Col(MG828)	-
PR256E1	South Africa	Rectal swab	SH005	88	*CTX-M-1*	-	-	-	-	*tet(A), sul2,*	IncI1[&], IncI2, Col(MG828), ColPVC, IncFIB,	IncF [K-:A-:B1]; IncI1[ST3]
PN256E2	South Africa	Nasal swab	SH005	88	*CTX-M-1*	-	-	-	-	*tet(A), sul2,*	IncI1*, IncFIB, Col(MG828), Col440I, rep10	IncF [K-:A-:B1]; IncI1[ST3]
PN027E1II	Cameroon	Nasal swab	SH001	226	*CTX-M-15*	*TEM-1B*	-	-	*QnrS1*	*aph(3'')-Ib, aph(6)-Id, tet(A), mdf(A), sul2, dfrA14,*	IncY, Col440I, colRNAI, Col(MG828)	-
PN091E1II	Cameroon	Nasal swab	SH002	940	*CTX-M-15*	*TEM-1B*	-	*gyrA (p.S83A),*	-	*aph(3'')-Ib, aph(6)-Id, aadA1, 16S_rrsC (g.926_926del), tet(B), mph(A), sul2, dfrA1,*	IncX, Col440I	-
PN256E8	South Africa	Nasal swab	SH005	944	*CTX-M-55*	*TEM-1B TEM-141 TEM-206*	-	-	*oqxA, oqxB aac(6')-Ib-cr*	*aac(6')-Ib3, aadA5, tet(A), sul2, dfrA17, floR, mcr-1.1, fosA3*	IncN, IncHI2A, IncHI2	IncN [ST1]; IncHI2 [ST3-like]
PR209E1	South Africa	Rectal swab	SH004	2144	*CTX-M-14*	-	-	-	*oqxB, oqxA*	*aph(3'')-Ib, aph(6)-Id, aadA2b, aadA1, sul3, cmlA1, fosA3*	IncFIC(FII), IncFIB, IncHI2A, IncHI2 rep21	IncF [K89:A-:B57] IncHI2[ST3]
PR246B1C	South Africa	Rectal swab	SH004	2144	*CTX-M-14*	-	-	-	*oqxA, oqxB*	*aph(3'')-Ib, aadA2b, aph(6)-Id, aadA1, aph(3'')-Ib, sul3, fosA3, cmlA1*	IncFIC(FII), Col440II, IncHI2A, IncHI2, IncFIB	IncF [K89:A-:B57] IncHI2 [ST3]
PR085E3	Cameroon	Rectal swab	SH002	4450	*CTX-M-15*	-	-	-	*QnrS1*	*AadA5, sul2, dfrA17*	IncY	-

Table 3. *In silico* identification of human pathogenicity and virulence factors in the ESBL-*E. coli* isolates

Pathogenicity feature	Nasal Isolates						Rectal isolates				
	PN017E2II	**PN027E6IIB**	**PN027E1II**	**PN091E1II**	**PN256E2**	**PN256E8**	**PR010E3I**	**PR209E1**	**PR246B1C**	**PR256E1**	**PR085E3**
Pathogenicity Score (No. of Pathogenic Families)	0.934 (615)	0.937 (889)	0.94 (526)	0.941 (665)	0.927 (735)	0.932 (625)	0.94 (677)	0.939 (710)	0.937 (682)	0.929 (729)	0.939 (666)
Human Pathogenicity	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Virulence factors											
Adherence	*ecpC, ecpD, ecpE, elfC, elfG, eaeH, hcpB, hcpC, fimD, fimH*	*lpfA,*	*elfC, elfG, hcpB, hcpC, fimD, fimH*	*cfaC, cfaD/cfaE, ecpC, ecpD, ecpE, elfC, elfG, eaeH, hcpB, hcpC, lpfA*	*hra, lpfA, air, tsh, cfaC, cfaD/cfaE, ecpC, ecpD, ecpE, elfC, elfG, eaeH, focD, focH, hcpB, hcpC, papC, papD, fimD, fimH, pilQ, pilR, pilV*	-	*hra, papA_F19, ecpA, ecpB, ecpC, ecpD, ecpE, elfA, elfC, elfD, elfG, eaeH, hcpA, hcpB, hcpC, papI, fimA, fimC, fimD, fimE, fimF, fimG, fimH,*	*lpfA, cfaA, cfaB, cfaD/cfaE, ecpA, ecpB, ecpC, ecpD, ecpE, elfA, elfC, elfD, elfG, eaeH, hcpA, hcpB, hcpC, papI, fimA, fimC, fimD, fimE, fimF, fimG, fimH, fimI*	*lpfA,*	*papC, focCsfaE, focF, focI, lpfA, tsh,*	*hra, lpfA, air, papC, tsh, cfaC, cfaD/cfaE, ecpC, ecpD, ecpE, elfC, elfG, eaeH, focC, hcpB, hcpC, papC, fimC, fimD, fimH, lfhA, prl/gapA, cgsG, pilW, staB, staC, stfC, stfD, stgB*
Autotransporter	*EhaB*	-	*aatA, ehaB, upaG/ehaG*	*ehaB, upaG/ehaG*	*agn43, ehaB, upaG/ehaG,*	-	*cah, ehaB*	*air/eaeX, ehaA, ehaB, upaG/ehaG,*	-	-	*cah, ehaA, ehaB, upaG/ehaG,*
Iron Uptake	-	*fyuA, irp2, sitA,*	-	*fyuA, irp1, irp2, ybtA, ybtE, ybtP, ybtS, ybtT, ybtU, ybtX*	*iucA, iucB, iucC, iucD, iutA, sitA, sitB, sitC, iroB, iroC, iroE, iroN, fyuA, irp1, irp2, ybtA, ybtE, ybtP, ybtS, ybtT, ybtU, ybtX*	-	*iutA, iucC, iucA, iucB, iucC, iucD, sitA, sitB, sitC, sitD*	-	-	*irp2, iucC, iutA, iroN, fyuA,*	*iroN, ccmF, entA, entB, entC, entE, entF, fepB, fepC, fepD, fepG, hemC, hemE, hemH, hemL, hemN*

Table 3. ***Continued***

Pathogenicity feature	Nasal Isolates						Rectal isolates				
	PN017E2II	**PN027E6IIB**	**PN027E1II**	**PN091E1II**	**PN256E2**	**PN256E8**	**PR010E3I**	**PR209E1**	**PR246B1C**	**PR256E1**	**PR085E3**
Secretion system	*aec15, aec18, aec22, aec23, aec25, aec26, aec27/clpV, aec29, aec30, aec31*	-	-	*aec15, aec18, aec22, aec23, aec25, aec26, aec27/clpV, aec29, aec30, aec31*	*aec15, aec18, aec22, aec23, aec25, aec26, aec27/clpV, aec28, aec29, aec30, aec31*	-	*aec15, aec18, aec22, aec23, aec25, aec26, aec27/clpV, aec28, aec29, aec30, aec31*	*aec15, aec18, aec22, aec23, aec25, aec26, aec27/clpV, aec28, aec29, aec30, aec31, aec32*	-	*etsC,*	*aec15, aec18, aec22, aec23, aec25, aec26, aec27/clpV, aec28, aec29, aec30, aec31, aec32, flgD, flgE, flgF, flgI, flgJ, flgK, flgL, flhA, flhB, fliA, fliF, flil, fliM, fliR, ipaH, gspC, gspD, gspE, gspF, gspK, gspL, clpB, rmkB, wbjD/wecB, wecC, galF, ugd, wcal, wzc*
Antiphagocytosis	*wzc, wzi*	-	-	*wzc, wzi, wbaZ*	-	-	*Wzi*	-	-	-	
Toxins	*hlyE/clyA*	-	-	-	*HlyF, astA, vat*	*astA*	*hlyE/clyA,*	*hlyE/clyA,*	-	*hlyF,*	*hlyE/clyA, hlyA*
Protectins and invasins	*ibeB, ibeC*	*KpsE, kpsMIII_K96; iss, ompT*	*ibeB, ibeC*	*iss ibeB, ibeC*	*KpsE, kpsMIII_K96, iss, ompT, traT, ibeB, ibeC*	*ompT,*	*traT, ibeB, ibeC, tia*	*iss, ompT, ibeB, ibeC,*	*traT, ompT, iss,*	*Iss, ompT, traT,*	*ibeB, ibeC, cheB, cheR, motA,*
Miscellaneous	*espL1, espL4, espX1, espY1, galE, rmlD, gad, terC*	*air, terC, gad, chuA, eilA,*	*espL1, espL4, espX1, espX4, espL1, espL4, espX1, espX5, rmlD,*	*espL1, espL4, espX4, terC, gad,*	*mchB, mchC, mchF, mcmA, terC, gad, eilA*	*terC, gad,*	*espL1, espL4, espX1, espX4, espX5, espY1, gad, terC, rmlD, galE, cea*	*espL1, espL4, espX1, espX4, espX5, terC, gad, adeG*	*gad, terC,*	*cea, cib, mchB, mchC, mcmA, terC,*	*espL1, espL4, espR1, espX5, galE, galU, mrsA/glmM, pgi, acpXL, rmlA, rmlD, rpoS, phoQ, glnA1, narH, sugC, acrB, farB, icl, mgtB, motB, bioB, katG, gmhA/lpcA, htrB, kdsA, kdtA, lpxA, lpxB, lpxK, msbA, opsX/rfaC, rfaD, rfaE, rfaF, wecA*

Table 4. Virulence-associated-traits of nasal and rectal ESBL-*E. coli*

Functional categories traits	Virulence factors	Overall, n=11 (%)	Nasal Isolates (n=5)	Rectal Isolates (n=6)
	Adhesins			
papACEFG	P fimbriae	4 (36)	1	3
fimH	Type 1 fimbriae	7 (64)	3	4
hra	Heat-resistant agglutinin	1 (9)	0	1
tsh	Temperature sensitive hemagglutinin	2 (18)	1	1
Sfa/foc	S or F1C fimbriae	1 (9)	0	1
	Toxins			
hlyD	α-hemolysin	1 (9)	0	1
hlyE	hemolysin E	4 (36)	1	3
hlyA	Hemolysin A	1 (9)	0	1
hlyF	putative avian hemolysin	2 (18)	1	1
astA	Enteroaggregative *E. coli* toxin	1 (9)	1	0
vat	Vacuolating toxin	1 (9)	1	0
	Siderophore			

iroN	Salmochelin receptor	2 (18)	1	1
fyuA	Yersiniabactin receptor	3 (27)	3	0
iutA	Aerobactin receptor	2 (18)	1	1
sitA	Periplasmic iron binding protein	3 (27)	2	1
	Protectins and Invasins			
kpsM_III	Group 3 capsule	2 (18)	2	0
kpsE	Group 2 capsule	2 (18)	2	0
iss	Increased serum survival	6 (55)	3	3
traT	Conjugal transfer surface exclusion protein	2 (18)	0	2
ompT	Outer membrane protease T	6 (55)	2	4
ibeB	Invasin of brain endothelial cells, IbeA	8 (73)	4	4
ibeC	Invasin of brain endothelial cells, IbeC	8 (73)	4	4
UpaG/ehaG	Autotransporter adhesin UpaG and EhaG	6 (55)	3	3
	Autotransporter proteins			
ehaB	Autotransporter adhesin	8 (73)	4	4

Table 5. Mobile genetic elements detected in the ESBL-*E. coli* isolates

Isolate (ST)	Plasmids	Insertion sequence	Transposons	Phages	CRISPR array (Cas system)	TR
PN017E2II (10)	IncY, Col(MG828), Col440I, rep21	-	-	-	6 (Cas1)	54
PR010E3I (44)	IncFIA, Col440I, IncFII, IncFIB, Col(MG828), rep21	-	-	-	8 (Cas1, Cas3)	48
PN027E6IIB (69)	IncY, Col(MG828)	ISKpn19, ISEc1, ISEc31, IS4, ISSfl10, IS911, cn_5813_IS911, MITEEc1, ISEc1, ISEc38, IS629, ISEc46, IS5075	-	PHAGE_Entero_mEp460_NC_019716	5 (Cas1)	54
PR256E1 (88)	IncI1[&], IncI2, Col(MG828), ColPVC, IncFIB,	IS26, ISVsa3, ISSbo1, cn_3792_ISSbo1, ISEc9, ISEc40, ISEc38, ISEc13	-	PHAGE_Entero_fiAA91_ss_NC_022750 PHAGE_Shigel_SfII_NC_021857(34) PHAGE_Entero_HK544_NC_019767	6 (Cas2)	51
PN256E2 (88)	IncI1*, IncFIB, Col(MG828), Col440I, rep10	IS26, ISVsa3, ISEc9	-	-	10 (Cas3)	101
PN027E1II (226)	IncY, Col440I, colRNAI, Col(MG828)	ISKpn19, ISEsa1, IS5075, MITEEc1, IS100, ISEc30, IS5, ISEc26, ISKpn8, IS421, IS609, ISEc38, IS30, IS903	-	-	11 (Cas3, Cas1)	55
PN091E1II (940)	IncX, Col440I	IS6100, MITEEc1, IS421, ISEc30, ISSfl10, IS30, ISEc38, ISEc1, IS100, ISKpn8	Tn7[#]	PHAGE_Entero_BP_4795_NC_004813	5 (Cas2)	40
PN256E8 (944)	IncN, IncHI2A, IncHI2	ISVsa3, IS640, IS100, ISEam1, IS30, MITEEc1, ISEc1, ISKpn26, IS421, ISVsa5, IS609	-	PHAGE_Shigel_SfII_NC_021857(34)	8 (Cas2)	87
PR209E1 (2144)	IncFIC(FII), IncFIB, IncHI2A, IncHI2 rep21	IS102, IS629, MITEEc1, ISKpn8, ISVsa5, IS421, IS3, IS26	Tn6082	PHAGE_Shigel_Sf6_NC_005344 PHAGE_Shigel_Sf6_NC_005344 PHAGE_Shigel_SfII_NC_021857(34)	6 (Cas2)	44
PR246B1C (2144)	IncFIC(FII), Col440II, IncHI2A, IncHI2, IncFIB	IS102, IS3, IS629, IS26, ISEc1, ISKpn8, ISVsa5, IS421, MITEEc1	Tn6082	PHAGE_Shigel_Sf6_NC_005344 PHAGE_Shigel_SfII_NC_021857 PHAGE_Entero_fiAA91_ss_NC_022750	8	41
PR085E3 (4450)	IncY	ISVsa3, ISEc9, IS421, ISKpn26, IS3, ISEc1, ISEc38, MITEEc1, IS26, IS102	-	PHAGE_Entero_mEp460_NC_019716 PHAGE_Pseudo_phiPSA1_NC_024365 PHAGE_Entero_fiAA91_ss_NC_022750	4 (Cas3)	39

TR: Tandem Repeat

Synteny of resistance and virulence genes and MGEs

[&]IncI1 (harbours 3 MGEs i.e IS26, ISVsa3, ISEc9 and encoded *sul2* and *cib*)

*IncI1 (harbours 3 MGEs i.e IS26, ISVsa3, ISEc9 and encoded *sul2* and *cib*)

[#]Tn7 (harbouring dfrA1)

Table 6. Virulence factors commonly present in ExPEC

Type	Associated virulence genes/proteins/enzyme
Toxins	hlyD (α-hemolysin), hlyA (hemolysin A), sat (secreted autotransporter toxin), pic (serine protease), vat (vacuolating toxin), astA (enteroaggregative *E. coli* toxin), cnf1 (cytotoxic necrotizing factor), cdt1 (cytolethal distending toxin), clb (colibactin)
Adhesins	papACEFG (P fimbriae), sfa/foc (S or F1C fimbriae), focG (F1C fimbriae adhesin), iha (adhesion siderophore), fimH (type 1 fimbriae), tsh (temperature sensitive hemagglutinin), hra (heat-resistant agglutinin), afa/draBC (D-binding adhesins), gaf (N-acetyl-D-glucosamine-specific fimbriae), bmaE (M fimbriae), iha (bifunctional enterobactin receptor/adhesion), ecpA (*E. coli* common pilus)
Siderophore	iroN (salmochelin receptor), fyuA (yersiniabactin receptor), ireA (siderophore receptor), iutA (aerobactin receptor), sitA (periplasmic iron binding protein)
Protectins and invasins	kpsM II (group 2 capsule), K1 (K1 group 2 capsule variants), K2 (K2 group 2 capsule variants), K5 (K5 group 2 capsule variants); KpsMT III (group 3 capsule), ibeA (invasion of brain endothelium), Cva (colicin V), traT (conjugual transfer surface exclusion protein), iss (increased serum survival), ompT (outer membrane protease T
Others	Usp (uropathogenic-specific protein), uidA (β-glucoronidase), H7 filC (flagellin variant), malX (pathogenicity island marker), dsdA (D-serine deaminase)

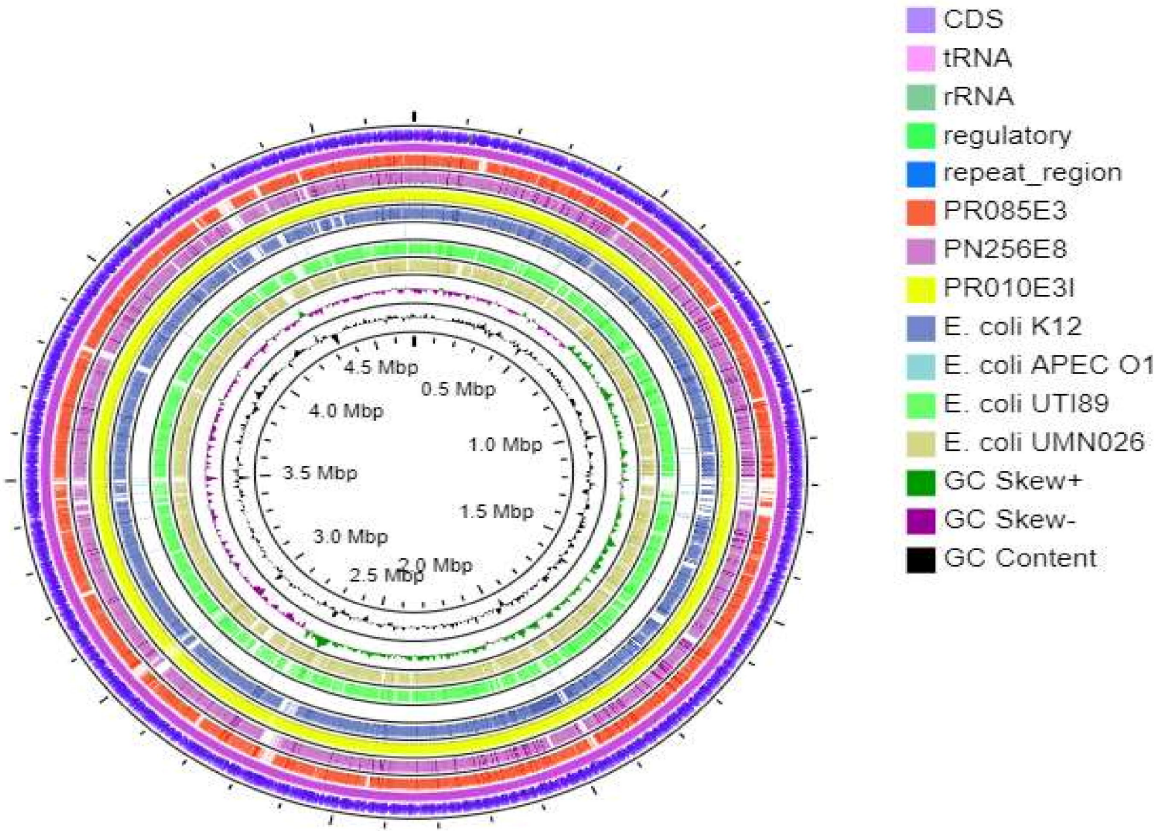

Figure 1. Circular genome representation of selected ESBL-*E. coli* aligned with reference genome and closely related strains. Circular map of selected ESBL-*E. coli* (PR010E3, PN256E8 and PR085E3) and comparative alignment with closely related strains (E. coli K12, E. coli APEC_O1, E. coli UTI89, E. coli UMN026), generated using CGView Server V1.0. Coloured arrows in the outer ring represent different gene families of the reference genome. A key of the coloured arrows representing different gene families is presented in the inset. The inner coloured circles representing different strains are also listed in the inset. Innermost circles show GC content indicated in black and GC Skew, with green and purple indicating positive and negative values, respectively.

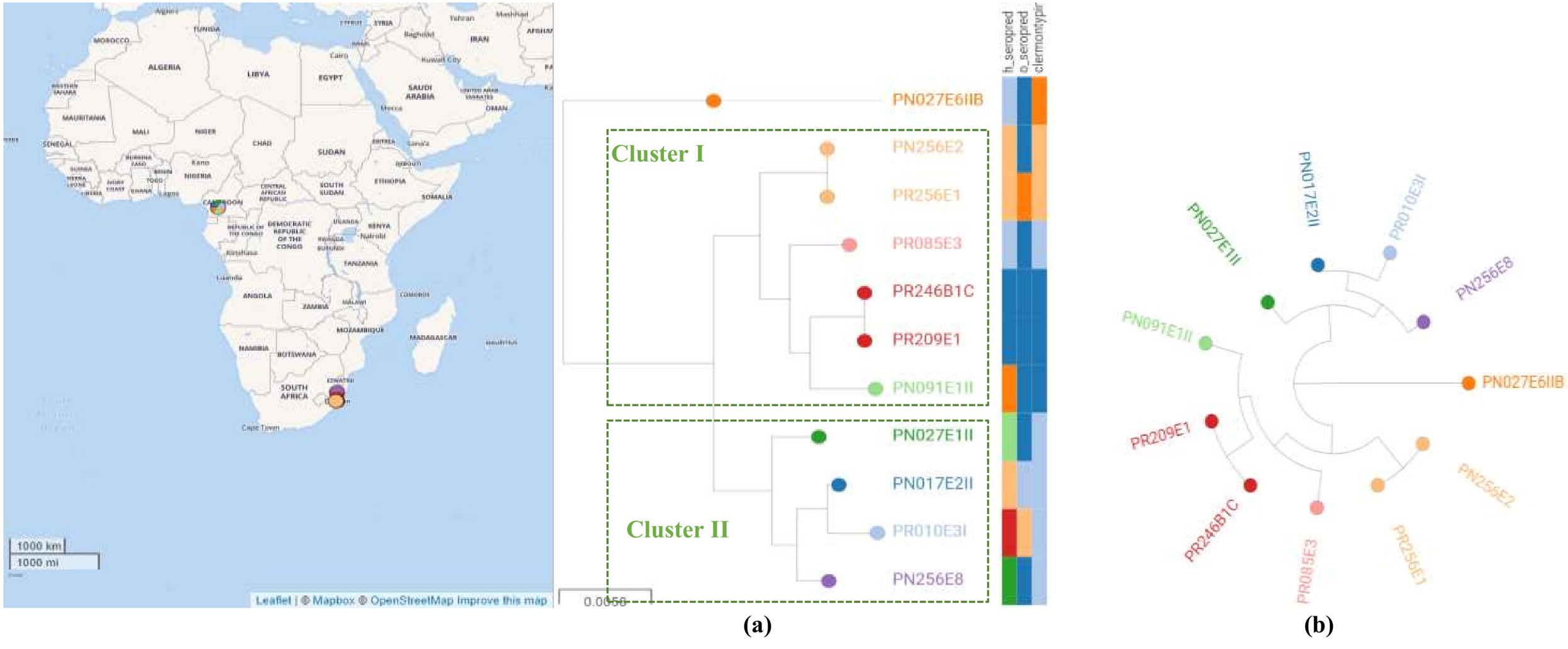

Figure 2. Comparative genome analysis based on the core genome MLST of *Escherichia coli* isolates. Each node represents an isolate, each of which is coloured according to its sequence type, as defined using the core genome. Clusters of isolates belonging to the same sequence cluster are encircled and annotated. Serotype and phylogroup are also indicated via a heatmap (a) Core-genome phylogenetic tree based on comparison of conserved clusters of orthologous genes (COGs). (b) Minimum spanning tree based on cgMLST. Interactive map of geographic locations and genetic attributes can be visualized within Microreact at https://microreact.org/project/tYENaUrCix7jMS7RBrFeBi/09dce9ac

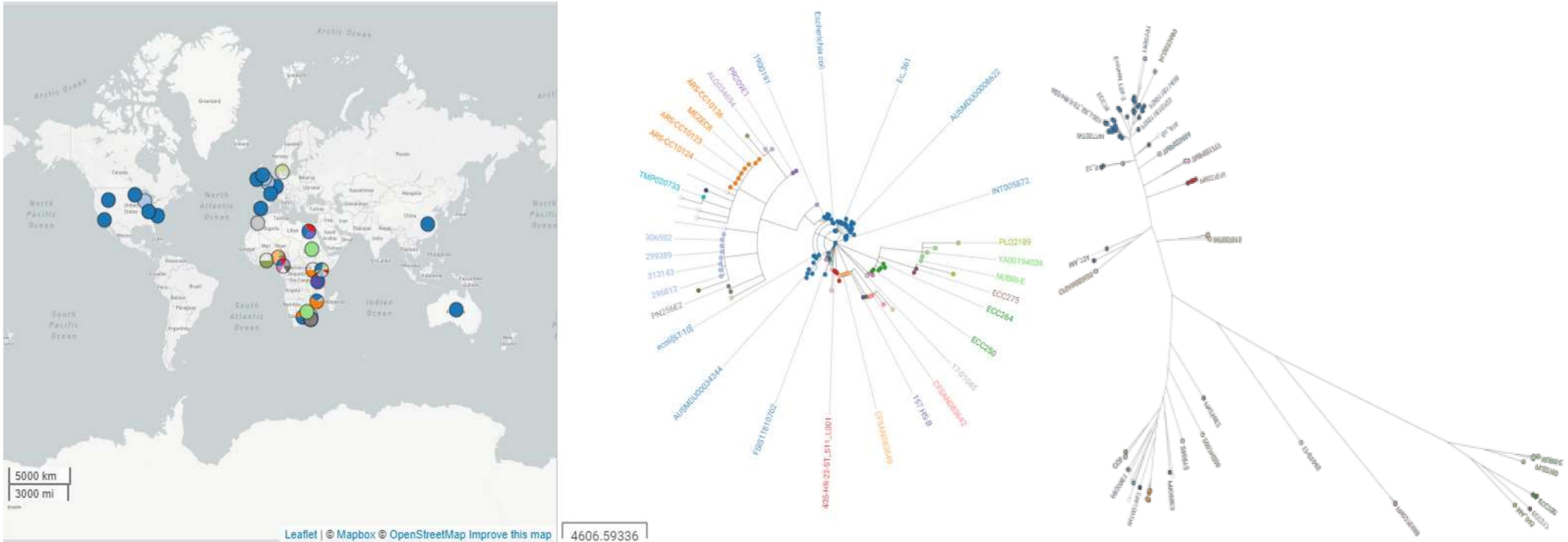

Figure 3. Comparative genome analysis based on the core genome MLST of the ESBL-*E. coli* with international *E. coli* isolates. Isolates of the same STs share the same colour. (a) Minimum spanning tree generated using cgMLST. (b) GrapeTree generated using MSTree V2 tool. Interactive map of geographic locations and genetic attributes can be visualized within Microreact at https://microreact.org/project/2kL6oNgm6VPKnUrrffkUD5/81a2cc99

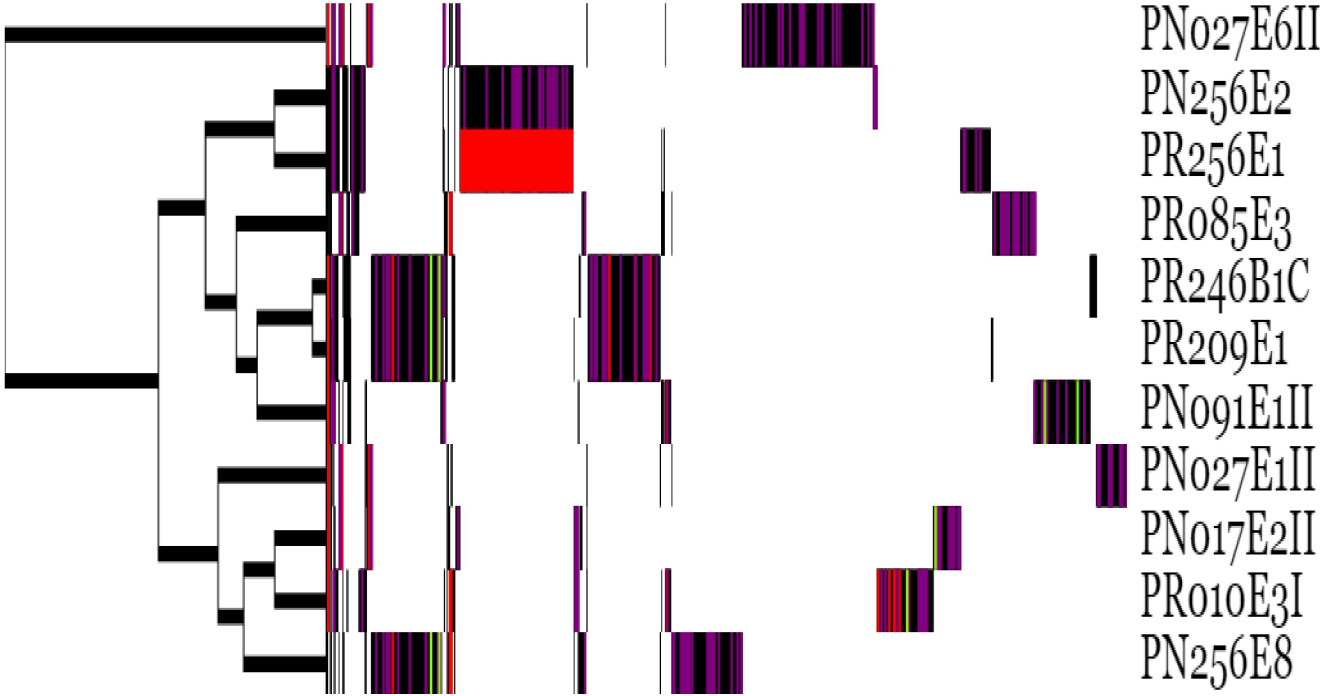

Figure 4. Comparative genome analysis based on the accessory genome of the ESBL-*E. coli* isolates. AMR genes are coloured in red, phages genes (VOGs) in purple and plasmids in are represented in green.