

Genomic epidemiology of *Escherichia coli* isolates from a tertiary referral center in Lilongwe, Malawi

Gerald Tegha^{1†}, Emily J. Ciccone^{2†}, Robert Krysiak¹, James Kaphatika¹, Tarsizio Chikaonda¹, Isaac Ndhlovu³, David van Duin², Irving Hoffman², Jonathan J. Juliano^{2,4,5,*}, ‡ and Jeremy Wang^{6‡}

Abstract

Antimicrobial resistance (AMR) is a global threat, including in sub-Saharan Africa. However, little is known about the genetics of resistant bacteria in the region. In Malawi, there is growing concern about increasing rates of antimicrobial resistance to most empirically used antimicrobials. The highly drug resistant *Escherichia coli* sequence type (ST) 131, which is associated with the extended spectrum β -lactamase *bla*_{CTX-M-15}, has been increasing in prevalence globally. Previous data from isolates collected between 2006 and 2013 in southern Malawi have revealed the presence of ST131 and the *bla*_{CTX-M-15} gene in the country. We performed whole genome sequencing (WGS) of 58 clinical *E. coli* isolates at Kamuzu Central Hospital, a tertiary care centre in central Malawi, collected from 2012 to 2018. We used Oxford Nanopore Technologies (ONT) sequencing, which was performed in Malawi. We show that ST131 is observed more often (14.9% increasing to 32.8%) and that the *bla*_{CTX-M-15} gene is occurring at a higher frequency (21.3% increasing to 44.8%). Phylogenetics indicates that isolates are highly related between the central and southern geographic regions and confirms that ST131 isolates are contained in a single group. All AMR genes, including *bla*_{CTX-M-15}, were widely distributed across sequence types. We also identified an increased number of ST410 isolates, which in this study tend to carry a plasmid-located copy of *bla*_{CTX-M-15} gene at a higher frequency than *bla*_{CTX-M-15} occurs in ST131. This study confirms the expanding nature of ST131 and the wide distribution of the *bla*_{CTX-M-15} gene in Malawi. We also highlight the feasibility of conducting longitudinal genomic epidemiology studies of important bacteria with the sequencing done on site using a nanopore platform that requires minimal infrastructure.

DATA SUMMARY

The sequencing data used for this analysis is available in public data repositories. Information on the newly generated data is available in NCBI SRA (BioProject ID PRJNA635644). Other publicly available sequences used are provided in Table S2.

INTRODUCTION

Antimicrobial resistance (AMR) is one of the most serious global public health threats [1]. Of specific concern are *Enterobacteriales*

(formerly the Enterobacteriaceae) that are resistant to third-generation cephalosporins such as ceftriaxone. The World Health Organization (WHO) has designated ceftriaxone-resistant *Enterobacteriales* as a critical priority [2]. In sub-Saharan Africa (SSA), there is growing evidence that ceftriaxone-resistant *Enterobacteriales* are important pathogens in invasive infections such as bacteremia. Given that ceftriaxone is often used to treat severe infections in SSA, and carbapenems are not often available, this is a major concern.

Received 29 May 2020; Accepted 17 November 2020; Published 09 December 2020

Author affiliations: ¹UNC Project, Lilongwe, Malawi; ²Division of Infectious Diseases, School of Medicine, University of North Carolina, USA; ³Kamuzu Central Hospital, Lilongwe, Malawi; ⁴Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina, USA; ⁵Curriculum in Genetics and Molecular Biology, School of Medicine, University of North Carolina, USA; ⁶Department of Genetics, School of Medicine, University of North Carolina, USA.

*Correspondence: Jonathan J. Juliano, jjuliano@med.unc.edu

Keywords: *Escherichia coli*; molecular epidemiology; whole genome sequencing; antimicrobial resistance; Africa; Malawi.

Abbreviations: AMR, antimicrobial resistance; CTX-M, cefotaxime-resistance Munich; ESBL, extended-spectrum β -lactamase; ExPEC, extraintestinal pathogenic *E. coli*; ONT, Oxford Nanopore Technology; SSA, sub-Saharan Africa; ST, sequence type; TMP-SMX, trimethoprim-sulfamethoxazole; WGS, whole genome sequencing.

All sequences generated for this project have been submitted to the NCBI SRA (BioProject ID PRJNA635644).

†These authors contributed equally to this work

‡These authors also contributed equally to this work

Data statement: All supporting data, code and protocols have been provided within the article or through supplementary data files. Six supplementary tables and two supplementary figures are available with the online version of this article.

000490 © 2021 The Authors



This is an open-access article distributed under the terms of the Creative Commons Attribution License.

Among the *Enterobacteriales*, *Escherichia coli* (*E. coli*) is a common cause of invasive disease, accounting for between 3 and 33 % of positive blood cultures in case series in Africa [3–8]. *E. coli* is becoming more resistant to commonly used antibiotics in SSA, including ceftriaxone. Additionally, recent evidence indicates that the highly drug resistant *E. coli* sequence type (ST) 131 has been increasing in prevalence globally [9–11]. This sequence type is an extraintestinal pathogenic *E. coli* (ExPEC) that is associated with blood-stream and urinary tract infections, often possessing genes associated with extended-spectrum β -lactamases (ESBLs) [12, 13]. The main mechanism of cephalosporin resistance is drug inactivation mediated by hydrolysis of the β -lactam ring by ESBL enzymes. Cefotaxime-resistance Munich (CTX-M) derivatives are the dominant and most widely distributed ESBL enzymes among *E. coli* [14]. CTX-M-15 is strongly associated with ST131 [15–18]. The global spread of ESBL-*E. coli* is largely attributed to the dissemination of *E. coli* strains carrying the *bla*_{CTX-M-15} gene, especially *E. coli* O25b:H4-ST131 [11]. Previously, three major lineages of ST131 have been identified that differed mainly with respect to their *fimH* alleles: A (mainly *fimH41*), B (mainly *fimH22*) and C (mainly *fimH30*) [19]. Clade C has predominated since the 2000s, corresponding with the rapid dissemination of the *bla*_{CTX-M-15} allele [11, 19, 20]. There is growing evidence in SSA that ST 131 CTX-M-15 *E. coli* strains are increasing, but there remains a limited number of studies assessing the clonality of *E. coli*, the distribution of ST131 and the presence of *bla*_{CTX-M-15} genes using whole genome sequencing [21–24].

In Malawi, specifically, there has been growing concern about increasing rates of antimicrobial resistance to most empirically used antimicrobials [25, 26]. A recent genomic epidemiology study of 94 *E. coli* isolates collected at Queen Elizabeth Central Hospital, a tertiary care centre in southern Malawi, and analysed by whole genome sequencing (WGS), has shown that ST131 is the most common ST in southern Malawi at 14.9% of isolates sequenced. CTX-M-15 was found in 21.4% of ST131 isolates, but occurred across 11 STs [27]. The purpose of our study is to increase our understanding of the genomic epidemiology of *E. coli* in Malawi by conducting WGS, using Oxford Nanopore Technologies (ONT) sequencing performed in Malawi, of isolates collected at a tertiary care hospital in central Malawi. We use this data to define the clonality, virulence genes and antimicrobial resistance genes in the central region of the country and to compare these results with those for southern Malawi.

METHODS

Sample selection

Sixty isolates were selected from the UNC Project-Malawi archives for sequencing. Samples were collected between 2012 and 2018. Organisms were isolated using sheep blood agar (trypticase soy agar prepared with 5% defibrinated sheep blood) and MacConkey agar (selective and differential medium for Gram-negative rods) as primary culture media. Depending on the year of sample collection, identification was

Impact Statement

Antimicrobial resistance is a global public health emergency. Although rates of resistance are high in Africa, little is known about the genetics and resistance mechanisms of clinically important bacteria. Here we characterize the molecular epidemiology of *Escherichia coli* isolates from a tertiary referral hospital in Malawi and compare these with historical isolates from the same country. Consistent with their global expansion, we show that ST131 is observed more often and that the *bla*_{CTX-M-15} gene is occurring at a higher frequency between studies. However, phylogenetics indicates that isolates are highly related between the central and southern geographic regions. This study highlights the feasibility of conducting longitudinal genomic epidemiology studies of important bacteria with the sequencing done on site using a nanopore platform that requires minimal infrastructure.

done either using conventional biochemical tests (TSI, indole, citrate and urease) and or Analytical Profile Index (API) 20E, a biochemical panel for identification and differentiation of members of the family Enterobacteriaceae. Isolates were selected for this study on the basis of diversity of phenotypic resistance pattern and clinical source of isolation, similarly to a previous study in Malawi [27]. One isolate was excluded from further analysis due to poor sequencing coverage, and one was excluded as it was identified as *Klebsiella*, resulting in 58 isolates included in all analyses.

Antimicrobial resistance testing

The Kirby-Bauer disc diffusion method was used to measure the *in vitro* susceptibility of bacteria to antimicrobial agents at the UNC Project-Malawi microbiology laboratory at the time of isolate collection. Results were obtained with disc diffusion tests that use the principle of standardized methodology and zone diameter measurements correlated with minimum inhibitory concentrations (MICs) with strains known to be susceptible, intermediate, and resistant to various antibiotics. All aspects of the procedure were standardized as recommended by the Clinical and Laboratory Standards Institute (CLSI) in the document ‘*Performance Standards for Antimicrobial Susceptibility Testing*’ [28].

Whole genome sequencing

Overnight cultures of the isolates were grown in 5 ml of LB broth at 37 °C. Cell pellets from the broth culture were recovered from 1.5 ml centrifuged at 10 000 g for 2 min. Cell pellets were resuspended in 100 μ l of nuclease-free water. DNA was extracted from the resuspended pellet using the Zymo Quick-DNA Microprep Kit (cell suspensions protocol) as per the manufacturer’s instructions (Zymo Research). The DNA was quantified using the dsDNA kit on a Qubit 2.0 (Thermo Fisher). Equal amounts of DNA (~100 ng) from each isolate were used for library preparation using the Rapid Barcoding

(RBK-004) as per the manufacturer's protocol (Oxford Nanopore Technologies). Pooled libraries of 12 isolates were run on a R9.4.1 flow cell on a MinION/MinIT for 24 h at UNC Project-Malawi. Each flow cell was washed once per protocol and a second set of 12 isolates were run for an additional 24 h or until all pores were exhausted.

De novo genome assembly

Base calling of fast5 files was done with Guppy (version 3.4.5) with the R9.4.1 'high accuracy' model [29]. Samples were demultiplexed and adapters/barcodes trimmed using a custom tool, depore (version 0.1; <https://github.com/txje/depore>). Reads for each isolate were *de novo* assembled using Flye (version 2.7) [30]. By Flye's design and empirically [31], chimeric reads should not affect assembly performance and are much less prevalent in rapid (transposase-based) as opposed to ligation-based library preparation [32]. Assemblies were then polished four times with racon (version 1.3.2) and underwent a final polish with medaka (version 0.6.2) (<http://github.com/nanoporetech/medaka>) [33, 34].

Typing of isolates

Each assembly was aligned using minimap2 to databases of known serotypes, fimH type, genomic and plasmid sequence types including plasmid incompatibility, virulence factors, and antimicrobial resistance genes, listed in Table S1 (available in the online version of this article) [35–40]. Matches were made for each assembly using similar cutoffs to those used by the Centre for Genomic Epidemiology (CGE; <https://cge.cbs.dtu.dk/services/data.php>) tools. Specifically, we required 60% of the feature to match at >90% sequence identity to serotype markers (fliC, wzx/wzy), fimH variant, virulence genes, antimicrobial resistance genes, and plasmid incompatibility group markers. For multi-locus sequence type (MLST) and fimH type, we report the closest hit (or multiple hits in case of a tie). Assembled contigs were identified as plasmids if they were found to contain at least one plasmid incompatibility type marker. For *bla*_{CTX-M-15} analysis, contigs >600 kbp without plasmid markers are inferred to be genomic, in all but two cases they are ~5 Mbp. No plasmid markers were found on any contig >300 kbp.

Species identification

We detected O and/or H serotypes and multi-locus sequence types in 59 of 60 presumed *E. coli* samples. The remaining sample appeared to have assembled well and a BLAST search revealed it was a strain of *Klebsiella*. We used this sequence as an outgroup in our phylogenetic analysis and otherwise excluded it from further analysis. An additional isolate (#31) was excluded from analysis for poor sequencing coverage, leaving 58 samples for all downstream analyses.

Phylogenomics

Assembled genomes were aligned to a set of single-copy orthologs largely conserved across *Enterobacterales* (BUSCO v4, <https://busco.ezlab.org/>). A subset of 92 genes were selected that appear in all 58 samples that were included in the

final analysis. A multiple-sequence alignment was performed for each gene using MUSCLE (3.8.31) [41]. This concatenated alignment includes 108 278 sites, 16 729 informative SNPs. RAXML-ng was used to reconstruct a maximum-likelihood phylogenetic tree for the concatenated alignments with model parameter 'GTR+G' (Fig. 1) [42]. For the comparison to existing Malawi *E. coli* isolates, we used publicly available sequence data listed in Table S2. For each of these samples, a genome was assembled from Illumina sequence data with SPAdes (3.14.0) using default parameters [43]. A phylogeny incorporating these sequences was generated as described above, using a subset of 80 genes present in all 149 genomes (Fig. 2). This 92 gene alignment includes 125 994 sites and 18 282 informative SNPs. To evaluate the accuracy of these phylogenies based on a limited set of highly conserved genes in representing the global genomic phylogeny, we generated a more liberal 'core' gene alignment using Roary [44] that captures a much larger portion of the genome and found few differences (Supplemental Methods, Figures S1 and S2).

Data analysis

Data were analysed using STATA/SE version 16.1 (StataCorp LLC, College Station, TX, USA). Descriptive statistics were used to describe clinical characteristics, frequencies of gene detections, and disc diffusion results. Given the small overall sample size, two-sided Fisher's exact tests were used to assess for associations between genes detected and phenotypic resistance testing by Kirby–Bauer disc diffusion. *P*-values were corrected for multiple comparisons using the Benjamini–Hochberg false discovery rate method [45].

RESULTS

Isolate characteristics and sequencing data

Five sequencing runs generated 9 291 492 reads totalling 40.8 Gbp with a mean read length of 4391 bp and N50 of 8732 bp. Sequencing summary statistics across samples are described in Table 1. Clinical and patient characteristics for the isolates are summarized in Table 2. The majority were from a urinary source (59%) and were from females (74%). A significant number of isolates were from sterile sites, including blood (24%), cerebrospinal fluid (3%), and joint (3%). Patient white blood cell counts and haemoglobin levels were available for 20 of the isolates; medians and interquartile ranges are included in Table 2.

Kirby–Bauer antimicrobial susceptibility testing results are summarized in Table 3. Notably, the majority of isolates were resistant to amoxicillin and trimethoprim–sulfamethoxazole (TMP–SMX) and 57 % (33/58) were resistant to ceftriaxone. These are commonly prescribed antibiotics in the outpatient (amoxicillin and TMP–SMX) and inpatient (ceftriaxone) settings in Malawi [46].

Sequence types, H and O groups

Twenty-one different ST groups were found among 53 isolates. Five isolates could not be assigned to a unique ST group (Table S3). ST131 was the most common type

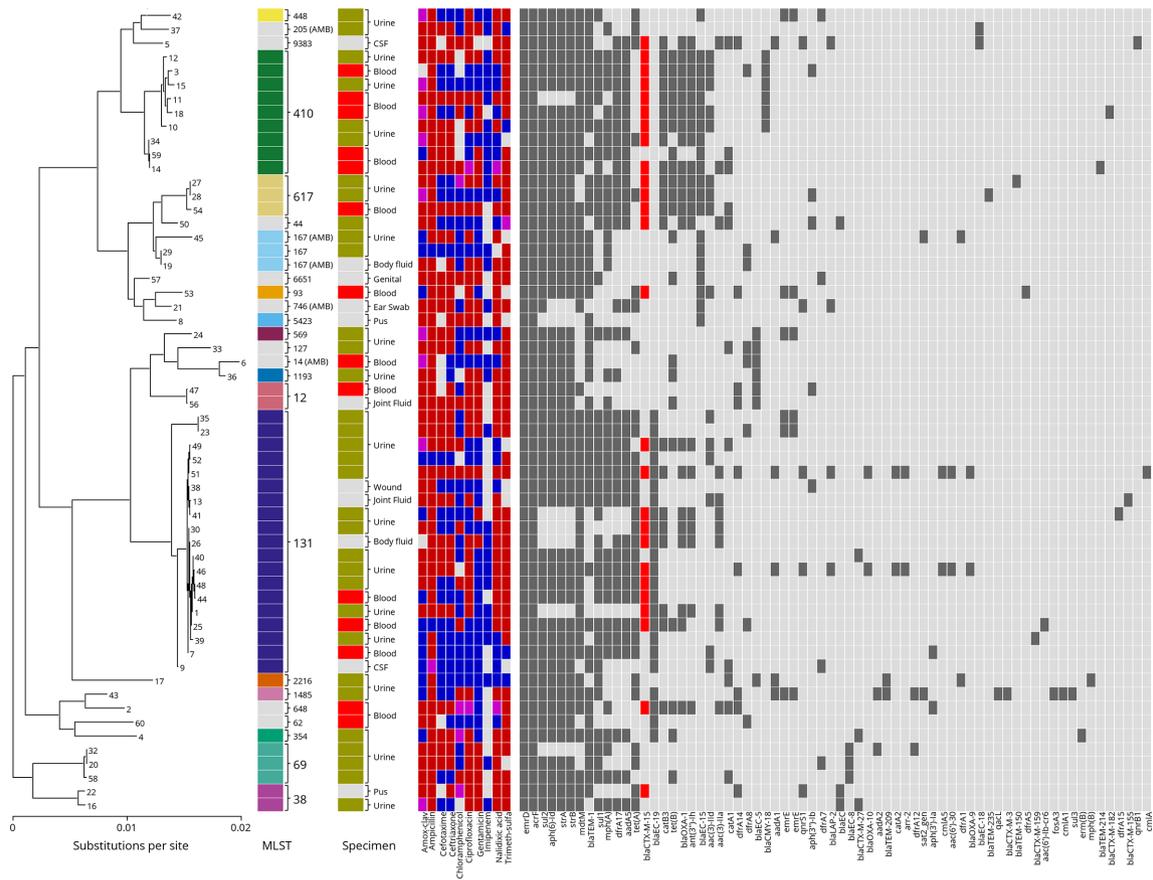


Fig. 1. Distribution of sequence type, specimen type, resistance phenotype and resistance gene composition in 58 Malawian *E. coli* isolates. Phylogenetic relationship among sequenced isolates with corresponding sequence type, specimen type, phenotypic and genomic AMR status. At the far left is the phylogeny relating these 58 samples with the sequenced *Klebsiella* from our study as an outgroup (not shown). In line with each terminal branch is the corresponding sample's sequence type (ST) ('AMB' indicates ambiguity in ST assignment), specimen from which the sample was isolated, AMR phenotype (red: resistant, blue: susceptible, purple: intermediate, grey: unknown), and presence of each detected AMR gene in the genome assembly (dark grey: present, light grey: absent, red highlights presence of $bla_{CTX-M-15}$).

identified [19 of 58 isolates (32.8%)], followed by ST410 which was present in 9 of 58 isolates (15.5%). Other STs that occurred in more than one isolate included ST69 (5.2%), ST38 (3.4%), ST617 (5.2%) and ST12 (3.4%). Of the 19 ST131, 18 of them were fimH30 and therefore clade C. One isolate was fimH27 and classified as clade B0. The fimH30 was linked to O25 and O16, while the fimH27 strain was O18 but also ST131. The population contained 23 different O groups (Table S4), with three samples being unable to be called, and 17 different H group calls (Table S5). A complete data set of genotype calls and sequencing characteristics is available in Table S6.

Population structure of *E. coli* in Malawi

A phylogenetic tree of the isolates with previously reported *E. coli* genomes from southern Malawi [27] is shown in Fig. 2. Notably, the ST131 isolates cluster into a single group and have relatively little genetic distance between them in the tree.

Genetic determinants of antimicrobial resistance

Our analysis identified 69 unique AMR genes that are known to encode proteins associated with antimicrobial susceptibility across a range of compounds (Table 4). AMR genes occurred across a range of ST and phylogenetic groups (Fig. 1).

β -Lactam and ESBL resistance

We identified 23 genes associated with β -lactam resistance, including 10 ESBL genes (Table 4). Extended-spectrum β -lactamases included $bla_{TEM-113}$, $bla_{CTX-M-15}$, $bla_{CTX-M-155}$, $bla_{CTX-M-159}$, $bla_{CTX-M-182}$, $bla_{CTX-M-27}$, $bla_{CTX-M-3}$, bla_{EC-15} , bla_{EC-18} and bla_{EC-19} . Forty-eight out of 58 (83%) isolates carried at least one ESBL gene, and a total of 40 (69%) isolates carried more than one ESBL gene. The presence of any ESBL gene was associated with phenotypic resistance to ceftriaxone ($P=0.004$). Fifteen isolates with an ESBL gene retained phenotypic susceptibility to ceftriaxone by disc diffusion. All of these isolates each had only one of the following ESBL genes: bla_{EC-15} , bla_{EC-18} , bla_{EC-19} .

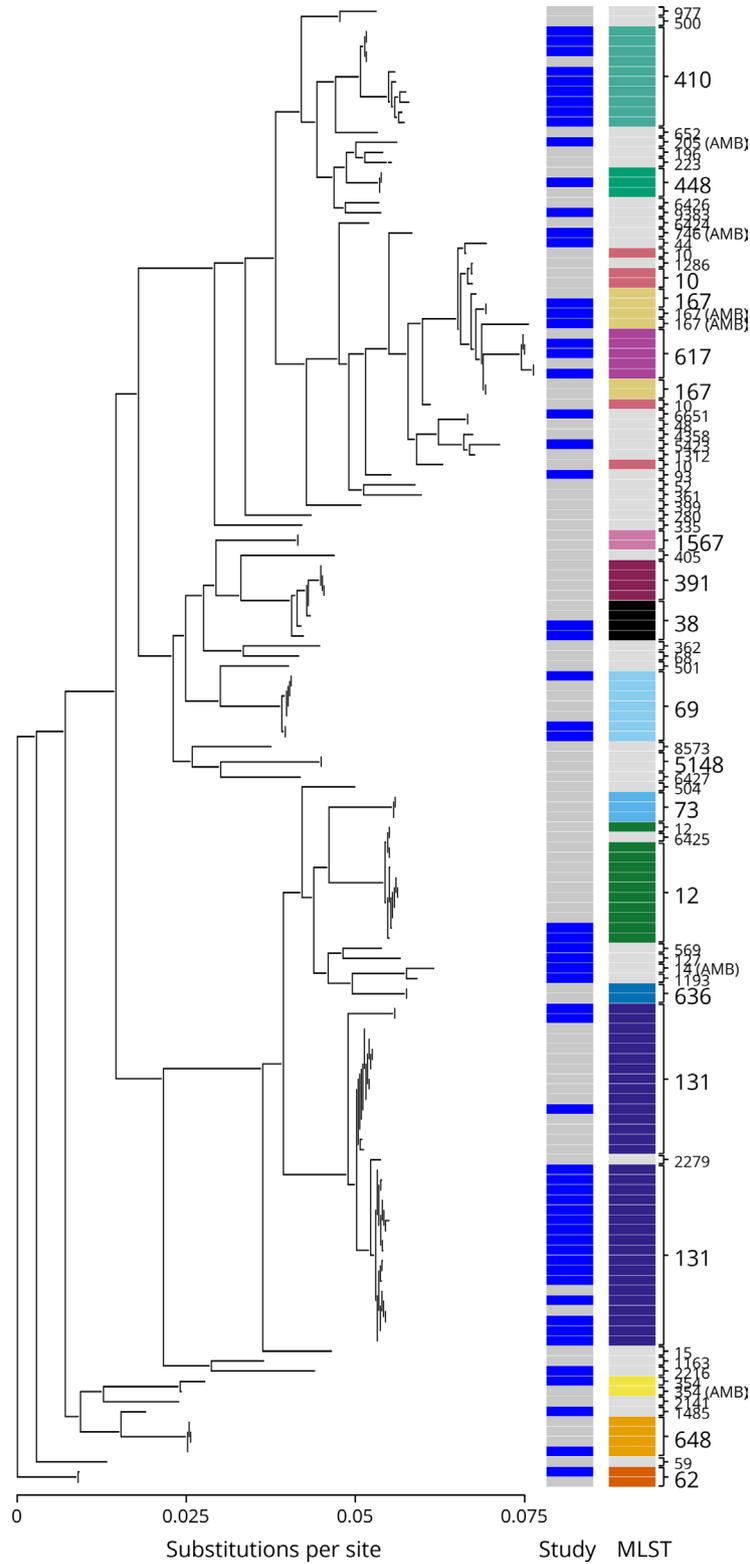


Fig. 2. Phylogenetic tree of Malawian *E. coli* isolates. Phylogenetic relationships among the 58 isolates are presented here in relation to previously published Malawi *E. coli* isolates (Table S2). Samples cluster by sequence type, but are well mixed between the two geographically and temporally separated studies. In the left column, blue indicates isolates from this study and grey from the previous study [17]. The rightmost column indicates the MLST where sequence types occurring more than once are assigned a unique colour (all others are left grey).

Table 1. Sequence data characteristics of all sequenced isolates

Sequence data	Value*
Number of reads	126 886 (21 843–539 714)
Total bps	556 Mbp (84–2076 Mbp)
Median fragment size sequenced in sample	2505 bp (673–4560 bp)
N ₅₀ assembly	4.88 Mbp (0.59–5.37 Mbp)
Number of contigs	5 (1–43)
Chromosomal median coverage	106 (15–392)

*Median (Minimum, Maximum).

The most common ESBL gene detected was *bla*_{CTX-M-15}, which occurred in 26 (44.8%) isolates (Table 5). Presence of the *bla*_{CTX-M-15} gene was associated with ceftriaxone resistance, with all 26 isolates being resistant to ceftriaxone by disc diffusion. The *bla*_{CTX-M-15} gene was found in isolates of eight different sequence types, most commonly ST131 [10 out of 19 isolates (52.6%)] and ST410 [8 out of 9 (88.9%) isolates]. We identified 15 O25b:H4-ST131 isolates, of which 10 (66%) had the *bla*_{CTX-M-15} gene. In ST131, the *bla*_{CTX-M-15} gene was highly associated with a chromosomal location with eight isolates (80%) having chromosomal copies, one isolate had a plasmid copy and one isolate with the gene in both locations. ST410 had a very different distribution with six of the eight isolates (75%) containing the *bla*_{CTX-M-15} gene in a plasmid-mediated

Table 2. Patient characteristics of the 58 included isolates

Variable	Number of isolates n (%)
Sex	
Female	43 (74)
Male	15 (26)
Age*	34 (20–47)
Sample Type	
Blood	14 (24)
Urine	34 (59)
CSF	2 (3)
Body Fluid	2 (3)
Joint	2 (3)
Other	4 (7)
Lab Values†	
White blood cell count (cells μl ⁻¹)	7.5 (3.8–9.1)
Haemoglobin (g dl ⁻¹)	9.8 (8.4–11.6)

*Median (IQR); data available for 54 patients.

†Median (IQR); data available for 20 patients.

Table 3. Kirby–Bauer antimicrobial resistance patterns of included isolates*

Drug (Total Tested)	Susceptible, n (%)	Intermediate, n (%)†	Resistant, n (%)
Amox-Clav (56)	15 (27)	9 (16)	32 (57)
Amoxicillin (58)	3 (5)	1 (2)	54 (93)
Cefotaxime (49)	21 (43)	0 (0)	28 (57)
Ceftriaxone (58)	25 (43)	0 (0)	33 (57)
Chloramphenicol (48)	29 (60)	4 (8)	15 (31)
Ciprofloxacin (58)	21 (36)	2 (3)	35 (60)
Gentamicin (57)	30 (53)	0 (0)	27 (47)
Imipenem (31)	31 (100)	0 (0)	0 (0)
Nalidixic Acid (57)	17 (30)	2 (4)	38 (67)
TMP-SMX (50)	4 (8)	1 (2)	45 (90)

*Percentage may not equal 100% due to rounding.

†Intermediate susceptibility by Kirby–Bauer test is defined based on a breakpoint that includes isolates with 'zone diameters within the intermediate range that approach usually attainable blood and tissue levels.' Response rates may be lower than for susceptible isolates; it implies clinical efficacy in sites where drugs physiologically concentrate or when a higher-than-normal dosage of a drug is used.

copy, while only a single isolate had a chromosomal copy and a single isolate had it in both locations.

Fluoroquinolone resistance

We extracted the sequence of the *gyrA* gene from each isolate and translated them into protein sequences. We examined these for the presence of *gyrA* mutations that have previously been described in Malawi [27]. We identified the S83L mutation in 24 isolates and the D87N mutation in 22 isolates. In all cases, the D87N mutation co-occurs with the S83L mutation; in only two cases was a single mutation (S83L) observed. In addition, we identified the *qnrB1* gene and the *qnrS1* gene in 1 (2%) and 5 (9%) of isolates, respectively. These genes did not co-occur in isolates with *gyr* mutations.

Aminoglycoside resistance

We identified 13 genes associated with aminoglycoside resistance in these isolates (Table 4). Several genes were associated with gentamicin resistance by disc diffusion, *aac(3)-IIa* ($P < 0.001$), *aac(3)-IIc* ($P < 0.001$), and *ant(3'')-Ih* ($P = 0.004$). We also detected *strA* or *strB* in combination in 50 of the isolates, as has been seen in southern Malawi previously [27]. When found together, these genes confer resistance to streptomycin, which is used in tuberculosis therapy in Africa [47].

Table 4. Prevalence of AMR genes identified

Gene	Resistance Phenotype	Isolates, n (%)
Aminoglycoside Resistance		
<i>aac(3)-IIa</i>	Gentamicin	11 (19)
<i>aac(3)-IIId</i>	Gentamicin	15 (26)
<i>aac(6')-30</i>	Aminoglycoside	2 (3)
<i>aac(6')-Ib-cr6</i>	Amikacin, kanamycin, tobramycin, quinolone	1 (2)
<i>aadA1</i>	Streptomycin	6 (10)
<i>aadA2</i>	Streptomycin	2 (3)
<i>aadA5</i>	Streptomycin	34 (59)
<i>ant(3'')-Ih</i>	Streptomycin/spectinomycin	20 (34)
<i>aph(3'')-Ib</i>	Streptomycin	5 (9)
<i>aph(3')-Ia</i>	Kanamycin	2 (3)
<i>aph(6)-Id</i>	Streptomycin	50 (86)
<i>strA</i>	Streptomycin	50 (86)
<i>strB</i>	Streptomycin	50 (86)
Multidrug Efflux Pumps		
<i>acrF</i>	Multidrug efflux	58 (100)
<i>emrD</i>	Multidrug efflux	58 (100)
<i>emrE</i>	Multidrug efflux	6 (10)
<i>mdtM</i>	Multidrug efflux	46 (79)
β-lactam Resistance		
<i>blaCMY-18</i>	Cephalosporin	6 (10)
<i>blaCTX-M-15</i>	Cephalosporin	26 (45)
<i>blaCTX-M-155</i>	Cephalosporin	1 (2)
<i>blaCTX-M-159</i>	Cephalosporin	1 (2)
<i>blaCTX-M-182</i>	Cephalosporin	1 (2)
<i>blaCTX-M-27</i>	Cephalosporin	3 (5)
<i>blaCTX-M-3</i>	Cephalosporin	1 (2)
<i>blaEC-15</i>	Cephalosporin	20 (34)
<i>blaEC-18</i>	Cephalosporin	2 (3)
<i>blaEC-19</i>	Cephalosporin	23 (40)
<i>blaEC-5</i>	Cephalosporin	7 (12)
<i>blaEC-8</i>	Cephalosporin	3 (5)
<i>blaEC</i>	β -lactam	3 (5)
<i>blaLAP-2</i>	β -lactam	4 (7)
<i>blaOXA-1</i>	Cephalosporin	21 (36)
<i>blaOXA-10</i>	Cephalosporin	2 (3)
<i>blaOXA-9</i>	β -lactam	2 (3)
<i>blaTEM-1</i>	β -lactam	43 (74)

Continued

Table 4. Continued

Gene	Resistance Phenotype	Isolates, n (%)
<i>blaTEM-113</i>	Cephalosporin	1 (2)
<i>blaTEM-150</i>	β -lactam	1 (2)
<i>blaTEM-209</i>	β -lactam	2 (3)
<i>blaTEM-214</i>	β -lactam	1 (2)
<i>blaTEM-235</i>	β -lactam	1 (2)
Chloramphenicol Resistance		
<i>catA1</i>	Chloramphenicol	11 (19)
<i>catA2</i>	Chloramphenicol	2 (3)
<i>catB3</i>	Chloramphenicol	21 (36)
<i>cmlA</i>	Chloramphenicol	1 (2)
<i>cmlA1</i>	Chloramphenicol	1 (2)
<i>cmlA5</i>	Chloramphenicol	2 (3)
TMP/SMX Resistance		
<i>dfrA1</i>	Trimethoprim	2 (3)
<i>dfrA12</i>	Trimethoprim	2 (3)
<i>dfrA14</i>	Trimethoprim	9 (16)
<i>dfrA15</i>	Trimethoprim	1 (2)
<i>dfrA17</i>	Trimethoprim	35 (60)
<i>dfrA5</i>	Trimethoprim	1 (2)
<i>dfrA7</i>	Trimethoprim	4 (7)
<i>dfrA8</i>	Trimethoprim	8 (14)
<i>sul1</i>	Sulfonamide	39 (67)
<i>sul2</i>	Sulfonamide	51 (88)
<i>sul3</i>	Sulfonamide	1 (2)
Quinolone Resistance		
<i>qnrB1</i>	Quinolone	1 (2)
<i>qnrS1</i>	Quinolone	5 (9)
Other Antimicrobials		
<i>arr-2</i>	Rifamycin	2 (3)
<i>aar-3</i>	Rifamycin	1 (2)
<i>erm(B)</i>	Macrolide	1 (2)
<i>fosA3</i>	Fosfomycin	1 (2)
<i>mph(A)</i>	Macrolide	37 (64)
<i>mph(B)</i>	Macrolide	1 (2)
<i>qacL</i>	Quaternary Ammonium	1 (2)
<i>sat2_gen</i>	Streptothricin	2 (3)
<i>tet(A)</i>	Tetracycline	30 (52)
<i>tet(B)</i>	Tetracycline	21 (36)

Table 5. Characteristics of CTX-M-15 associated isolates and ceftriaxone disk diffusion results

Isolate	Year Collected	Source	Sequence Type (ST)	Plasmid Inclusion Group(s)	Disk Diffusion Result	Location of Gene
1	2018	Urine	131	incB/O/K/Z	R	Genomic
2	2017	Blood	648	incF	R	Genomic
3	2018	Urine	410	incF	R	Plasmid
5	2018	Blood	9398	incHI2, incY	R	Plasmid
10	2018	Body Fluid	410	incF	R	Plasmid
11	2017	Urine	410	incF	R	Plasmid
12	2017	Urine	410	incF	R	Both
14	2014	Pus	410	incF, incY	R	Plasmid
15	2018	Urine	410	incF	R	Plasmid
18	2018	Body Fluid	410	incF	R	Genomic
22	2018	Blood	38	incF	R	Plasmid
25	2014	Urine	131	incF	R	Genomic
26	2014	Urine	131	incF	R	Plasmid
27	2015	Urine	617	incF	R	Both
28	2014	Urine	617	incF, incN	R	Plasmid
30	2016	Urine	131	incF	R	Genomic
34	2016	Urine	410	incF	R	Plasmid
41	2013	Blood	131	incN	R	Genomic
44	2013	Urine	131	incF	R	Genomic
46	2013	Urine	131	incF	R	Genomic
48	2014	Blood	131	incF	R	Genomic
49	2013	Blood	131	incF	R	Genomic
50	2013	Joint Fluid	44	incF	R	Plasmid
51	2013	Genital Swab	131	incF	R	Both
53	2015	Blood	93	incI-1I, incF	R	Plasmid
54	2013	Cerebrospinal Fluid	617	incF	R	Plasmid

Resistance to other antimicrobials

We identified six genes associated with chloramphenicol resistance, the most common being *catA1* in 11 (19%) of isolates and *catB3* in 21 (36%) of isolates. We did not identify isolates with *floR*, which has previously been identified in southern Malawi [27]. Detection of *catA1* was associated with intermediate susceptibility or resistance to chloramphenicol by disc diffusion ($P=0.004$), but detection of *catB3* was not ($P=0.613$). We identified 11 genes associated with resistance to TMP-SMX, the most common being the trimethoprim resistance gene *dfrA17* (60% of isolates) and the sulfonamide resistance genes *sul1* (67% of isolates) and *sul2* (88% of isolates). Of these three genes, only *sul2* detection was associated with TMP-SMX intermediate susceptibility or resistance by disc diffusion ($P=0.025$); all isolates that were

phenotypically intermediate or resistant to TMP-SMX had *sul1* and/or *sul2*. We detected *sul1* and *sul2* in 35 of the 58 isolates. Finally, we identified a handful of resistance genes to other antimicrobials, including rifamycin, macrolides, fosfomycin, and tetracyclines (Table 4). Interestingly, the most common AMR genes detected were multidrug efflux pumps, *acrF* and *emrD*, both of which occurred in all isolates.

Plasmid incompatibility group

We identified eight different plasmid incompatibility groups (Table 6). The most common were incF incompatibility groups, at least one variant of which (incFIA/incFIB/incFII) was found in all but seven isolates (88%). Although the majority of ST131 were in this group, it was not associated with ST131 ($P=0.567$) as many non-ST131 isolates had incF

Table 6. Prevalence of plasmid incompatibility groups

Plasmid Incompatibility Group	Number of isolates, n (%)
incFII	26 (45)
incFIA	19 (33)
incFIB	15 (26)
incY	4 (7)
incB/O/K/Z	2 (3)
incI	2 (3)
inc11-I	2 (3)
incN	2 (3)

incompatibility groups as well. Please see Table S6 for all inc type combinations.

Virulence genes

In total, we identified 37 virulence genes (Table 7) with each isolate carrying a median of seven genes (IQR 3–10). The virulence factors spanned a range of functions, including acid resistance, adhesion, invasins, metalloproteases, and toxins. The most common gene detected was *gad*, which occurred in all 58 isolates and encodes glutamate decarboxylase, an enzyme linked with bacterial ability to resist environmental stresses [48]. Multiple adhesin proteins were also identified. The most common of these were *papC*, *papH*, *papG-II*, and *lpfA*. *papC* (36% of isolates), *papH* (35% of isolates), *papG-II* (35% of isolates) are all involved in pili function. *lpfA* encodes the long polar fimbriae associated with human diarrheal disease, and occurred in 19 of 58 isolates [49]. A single protectin encoded by *iss* was identified and was the second most common virulence factor identified in 35 of 58 isolates. The *iss* (increased serum survival) gene was first identified in a human septicemic *E. coli* isolate and was associated with a 20-fold increase in complement resistance and a 100-fold increase in virulence toward 1-day-old chicks [50–52]. Multiple siderophores were identified, with the most common being *iha*, occurring in 29 of 58 isolates [53]. We also identified two common toxins, *sat* which occurred in 23 of 58 isolates and *senB* in 22 of 58 isolates. The secreted autotransporter toxin (*sat*) appears to fall within one subgroup of autotransporters recently classified as the serine protease autotransporters of *Enterobacteriaceae* (SPATE) family. It acts as a vacuolating cytotoxin for bladder and kidney epithelial cells [54]. The *senB* gene encodes the TieB protein, which may have some role in enterotoxicity of EIEC [55, 56].

DISCUSSION

In this study, we sequenced 58 *E. coli* isolates collected at a tertiary care centre in Lilongwe, Malawi between 2012–2018. To our knowledge, this is one of the first studies from Malawi demonstrating the feasibility of performing all steps of the whole genome sequencing process, including DNA extraction,

Table 7. Prevalence of virulence genes identified

Category	Gene	Isolate, n (%)	
Acid resistance	<i>gad*</i>	58 (100)	
Adhesin	<i>afaC</i>	13 (22)	
	<i>air</i>	9 (16)	
	<i>lpfA</i>	19 (33)	
	<i>nfaE</i>	13 (22)	
	<i>papA</i>	1 (2)	
	<i>papC</i>	21 (36)	
	<i>papE</i>	2 (3)	
	<i>papG-II</i>	20 (35)	
	<i>papG-III</i>	1 (2)	
	<i>papH</i>	20 (35)	
	<i>sfaF</i>	3 (5)	
	<i>sfaS</i>	1 (2)	
Effector (T3SS)	<i>espl</i>	1 (2)	
Invasin	<i>ibeA</i>	1 (2)	
Metalloprotease	<i>sslE</i>	23 (40)	
Microcin	<i>mchB</i>	2 (3)	
	<i>mchC</i>	3 (5)	
	<i>mchF</i>	6 (10)	
	<i>mcmA</i>	3 (5)	
Protectin	<i>iss</i>	35 (60)	
Regulator	<i>eilA</i>	9 (16)	
Siderophore	<i>iha</i>	29 (50)	
	<i>ireA</i>	6 (10)	
	<i>iroN</i>	8 (13)	
	<i>iutA</i>	3 (5)	
Toxin	<i>cma</i>	1 (2)	
	<i>cnf1</i>	6 (10)	
	<i>hly-alpha</i>	9 (16)	
	<i>pic</i>	1 (2)	
	<i>sat</i>	23 (40)	
	<i>senB</i>	22 (38)	
	<i>tsh</i>	1 (2)	
Transferase	<i>vat</i>	6 (10)	
	<i>capU</i>	7 (12)	
ATP-binding cassette transporter	<i>ybtP</i>	20 (35)	
	<i>ybtQ</i>	22 (38)	

*This gene is almost universal to *E. coli* and can be found in non-pathogenic strains.

library preparation, and the sequencing itself, on site in a local laboratory.

Relatively little data exist concerning the genomics of *E. coli* in Africa to date [21–24]. The isolates sequenced as part of our project collected in Lilongwe, in the central region of Malawi, are highly phylogenetically similar to those seen in a previous study conducted in Blantyre, Malawi, in the southern region [27]. Similarly to the previous report, we identified a diverse set of AMR genes, with similar genes occurring across a range of *E. coli* lineages in Malawi. We also provide additional information on common virulence genes associated with *E. coli* in Malawi, many of which are shared across a range of lineages.

In this joint phylogenetic analysis, we combined nanopore-only assembled genomes with solely short-read-based assemblies. This approach has the potential to introduce biases in both gene prediction and single-nucleotide polymorphisms due to the divergent error profiles of the two technologies. Nanopore sequencing in particular is known to suffer from a relatively high error rate among raw reads that is dominated by short insertions and deletions and primarily in homopolymer stretches [57]. Although we sequenced to sufficient depth that the vast majority of these errors were eliminated during the consensus generation and multiple polishing steps that produced the final assembly, undoubtedly there existed some residual errors above the rate typically observed with short-read sequencing platforms like Illumina [58]. Where these platform-specific errors are shared among samples sequenced using the same platform, there is the possibility of these segregating biases making their way into the predicted phylogeny. While we observe appropriate clustering of clades by sequence type regardless of sequencing approach (Fig. 2), there is, in several cases, apparent segregation of samples within STs (notably ST131) that is consistent with either true phylogenetic structure or segregating sequencing platform-specific biases. Since the samples collected in our study and the previous study [27] are separated both temporally and geographically, it is reasonable to expect within-ST variation to be attributable to segregating biological variation, but we cannot conclusively rule out that sequencing bias contributes to this phenomenon without careful manual inspection of the segregating sites.

Consistent with the Blantyre study's findings [27], we show that *E. coli* is highly diverse, with a distribution of STs similar to global isolates. Importantly, the previous study utilized samples collected between 1996 and 2014, allowing temporal comparisons between studies as the majority of samples in our analysis were collected after 2013. Our results indicate that ST 131 has become more prevalent (14.9% increasing to 32.8%) and that the *bla*_{CTX-M-15} gene is occurring at a higher frequency (21.3% increasing to 44.8%) in the intervening years. This is consistent with global trends that indicate that the highly drug-resistant *E. coli* ST131, associated with the *bla*_{CTX-M-15} gene, has been increasing in prevalence [9–11, 15–18]. As expected, there is some subtle structure within ST131, with isolates from this study being primarily localized on a branch with a longer internal branch length (Fig. 2).

Another difference from previous work is the higher number of ST410 isolates in this study, which carried the *bla*_{CTX-M-15} gene at the highest frequency of all the sequence types. Overall, there does not appear to be any strong associations between overall pattern of AMR gene content, virulence gene content, isolation site, and ST within the isolates (Fig. 1), similarly to previous results [27].

Although globally there is a strong association between ST131 and presence of the *bla*_{CTX-M-15} gene, here we identified the *bla*_{CTX-M-15} gene across a diverse set of lineages [11]. This finding is similar to those from other studies in Tanzania and Malawi, where the *bla*_{CTX-M-15} gene was found across numerous STs [22, 27]. We see the *bla*_{CTX-M-15} gene in eight different sequence types, including ST131. The ST that most commonly contained the *bla*_{CTX-M-15} gene was ST410. All of the ST131 isolates that contained the *bla*_{CTX-M-15} gene were O25b:H4-ST131 in this study. Overwhelmingly, the ST131 isolates were clade C, containing the fimH30 gene, consistent with the global expansion of ST131-H30 [12]. Interestingly, there is a strong association between ST and where the *bla*_{CTX-M-15} gene is carried in the ST131 and ST410 isolates. ST131 was much more likely to have a genomic location for the gene, whereas ST410 more frequently carried the gene on a plasmid. This may have implications for how the *bla*_{CTX-M-15} gene spreads in Malawi. Given the diverse lineages that carry the *bla*_{CTX-M-15} gene, additional studies are needed to better understand the epidemiology of this gene in Malawi.

Overall, the patterns of AMR gene prevalence were similar to the one previous report from southern Malawi. *sul2*, *strA*, *strB*, *dfrA*, *bla*_{TEM-1}, and *sul1* genes remained very common in the population. Interestingly, chloramphenicol resistance gene prevalence was lower than previously reported, with a decrease in the prevalence of *catA* from 64.9 % to 22 % [27]. This potentially reflects decreased use of the drug in the community with changing treatment guidelines and increasing availability of alternative agents with fewer side effects [46]. Most resistance genes were detected broadly across different genotypes in this study (Fig. 1).

In summary, we confirm that the *E. coli* population in Malawi is highly diverse, with evidence for increasing prevalence of the ST131 group in the country. We see a higher proportion of ST 131 isolates and a higher prevalence of the *bla*_{CTX-M-15} gene in our isolates, which were collected a few years later than those described in previous reports. This expansion is consistent with the global increase in O25b:H4-ST131 bearing the fimH30 gene. *E. coli* genotypes are similar between two major tertiary care hospitals that are quite distant, with highly related isolates being found between the sites. As previously seen, AMR genes, including the *bla*_{CTX-M-15} gene, are broadly contained across sequence types. A high diversity of virulence genes were seen within *E. coli* isolates. These data were collected by conducting ONT sequencing in Malawi, highlighting the possibility of conducting rapid longitudinal genomic epidemiology studies of consequential bacteria in sub-Saharan Africa where the sequencing is conducted on site.

Funding information

This project was funded by a Yang Biomedical Scholars Award from the University of North Carolina.

Acknowledgements

We would like to thank the laboratory team and management at UNC Project-Malawi for hosting the project. Thank you to the unknown individuals for use of their isolates.

Author contributions

G. T., contributed to investigation, data curation, and writing-review and editing. E. J. C., contributed to investigation, data curation, formal analysis, writing/original draft preparation, and project administration. R. K., contributed to investigation, resources and writing-review and editing. J. K., T. C. and I. N., contributed to investigation and data curation. D. v. D., contributed to writing-review and editing. I. H., contributed resources, writing-review and editing, and conceptualization. J. J. J., contributed conceptualization, methodology, formal analysis, resources, data curation, writing/original draft preparation, supervision, project administration, and funding. J. W., contributed conceptualization, methodology, software, formal analysis, resources, data curation, writing/original draft preparation, supervision, visualization, and project administration.

Conflicts of interest

D. v. D. has received funding or support from Allergan, Achaogen, Qpex, Shionogi, Tetrphase, Sanofi-Pasteur, T2 Biosystems, NeuMedicine, Roche, MedImmune, and Astellas. He also sits on a Merck Advisory Board. He has received travel reimbursement from the Infectious Disease Society of America (IDSA), American Society of Microbiology (ASM) and the European Society of Clinical Microbiology and Infectious Diseases (ESCMID). The other authors have no conflicts of interest to declare.

Ethical statement

Secondary use of stored clinical isolates from UNC Project Malawi was approved by the UNC IRB (18–3292) and National Health Sciences Research Committee IRB (approval number 2205). This work was deemed non-human subjects research.

References

- Prestinaci F, Pezzotti P, Pantosti A. Antimicrobial resistance: a global multifaceted phenomenon. *Pathog Glob Health* 2015;109:309–318.
- WHO. WHO publishes list of bacteria for which new antibiotics are urgently needed. <http://www.who.int/mediacentre/news/releases/2017/bacteria-antibiotics-needed/en/> (accessed 3 March 2020).
- Hill PC, Onyeama CO, Ikumapayi UNA, Secka O, Ameyaw S et al. Bacteraemia in patients admitted to an urban hospital in West Africa. *BMC Infect Dis*. 2007;7
- Okomo UA, Garba D, Fombah AE, Secka O, Ikumapayi UNA et al. Bacterial isolates and antibiotic sensitivity among Gambian children with severe acute malnutrition. *Int J Pediatr* 2011;2011:825123–.
- Nielsen MV, Sarpong N, Krumkamp R, Dekker D, Loag W et al. Incidence and characteristics of bacteremia among children in rural Ghana. *PLoS One* 2012;7:e44063.
- Enweronu-Laryea CC, Newman MJ. Changing pattern of bacterial isolates and antimicrobial susceptibility in neonatal infections in Korle BU teaching Hospital, Ghana. *East Afr Med J* 2007;84:136–140.
- Samuel SO, Fadeyi A, Akanbi AA, Ameen NB, Nwabuisi C et al. Bacterial isolates of blood cultures in patients with suspected septicemia in Ilorin, Nigeria. *Afr J Med Med Sci* 2006;35:137–141.
- Popoola O, Kehinde A, Ogunleye V, Adewusi OJ, Toy T et al. Bacteremia among febrile patients attending selected healthcare facilities in Ibadan, Nigeria. *Clin Infect Dis* 2019;69:S466–S473.
- Poolman JT, Wacker M. Extraintestinal pathogenic *Escherichia coli*, a common human pathogen: challenges for vaccine development and progress in the field. *J Infect Dis* 2016;213:6–13.
- Decano AG, Downing T. An *Escherichia coli* ST131 pangenome atlas reveals population structure and evolution across 4,071 isolates. *Sci Rep* 2019;9:17394.
- Ludden C, Decano AG, Jamrozny D, Pickard D, Morris D et al. Genomic surveillance of *Escherichia coli* ST131 identifies local expansion and serial replacement of subclones. *Microb Genom* 2020;6.
- Banerjee R, Johnson JR. A new clone sweeps clean: the enigmatic emergence of *Escherichia coli* sequence type 131. *Antimicrob Agents Chemother* 2014;58:4997–5004.
- Findlay J, Gould VC, North P, Bowker KE, Martin Williams O et al. Characterisation of cefotaxime-resistant urinary *Escherichia coli* from primary care in south-west England:2017–2018.
- Chaudhuri RR, Henderson IR. The evolution of the *Escherichia coli* phylogeny. *Infect Genet Evol* 2012;12:214–226.
- D'Andrea MM, Arena F, Pallecchi L, Rossolini GM. CTX-M-Type β -lactamases: a successful story of antibiotic resistance. *Int J Med Microbiol* 2013;303:305–317.
- Peirano G, Richardson D, Nigrin J, McGeer A, Loo V et al. High prevalence of ST131 isolates producing CTX-M-15 and CTX-M-14 among extended-spectrum- β -lactamase-producing *Escherichia coli* isolates from Canada. *Antimicrob Agents Chemother* 2010;54:1327–1330.
- Peirano G, Costello M, Pitout JDD. Molecular characteristics of extended-spectrum β -lactamase-producing *Escherichia coli* from the Chicago area: high prevalence of ST131 producing CTX-M-15 in community hospitals. *Int J Antimicrob Agents* 2010;36:19–23.
- Kanamori H, Parobek CM, Juliano JJ, Johnson JR, Johnston BD et al. Genomic analysis of multidrug-resistant *Escherichia coli* from North Carolina community hospitals: ongoing circulation of CTX-M-Producing ST131-H30Rx and ST131-H30R1 strains. *Antimicrob Agents Chemother* 2017;61 [Epub ahead of print 25 07 2017].
- Matsumura Y, Pitout JDD, Peirano G, DeVinney R, Noguchi T et al. Rapid identification of different *Escherichia coli* sequence type 131 clades. *Antimicrob Agents Chemother* 2017;61 [Epub ahead of print 25 07 2017].
- Cantón R, González-Alba JM, Galán JC. CTX-M enzymes: origin and diffusion. *Front Microbiol* 2012;3:110.
- Irengbe LM, Ambrose J, Bearzatto B, Durant J-F, Chirimwami RB et al. Whole-genome sequences of multidrug-resistant *Escherichia coli* in South-Kivu Province, Democratic Republic of Congo: characterization of phylogenomic changes, virulence and resistance genes. *BMC Infect Dis* 2019;19:137.
- Sonda T, Kumburu H, van Zwetselaar M, Alifrangis M, Mmbaga BT et al. Whole genome sequencing reveals high clonal diversity of *Escherichia coli* isolated from patients in a tertiary care hospital in Moshi, Tanzania. *Antimicrob Resist Infect Control*. 2018;7:72
- Ingle DJ, Levine MM, Kotloff KL, Holt KE, Robins-Browne RM. Dynamics of antimicrobial resistance in intestinal *Escherichia coli* from children in community settings in South Asia and sub-Saharan Africa. *Nat Microbiol* 2018;3:1063–1073.
- Mbelle NM, Feldman C, Osei Sekyere J, Maningi NE, Modipane L et al. The resistome, mobilome, virulome and phylogenomics of multidrug-resistant *Escherichia coli* clinical isolates from Pretoria, South Africa. *Sci Rep* 2019;9:16457.
- Iroh Tam P-Y, Musicha P, Kawaza K, Cornick J, Denis B et al. Emerging resistance to empiric antimicrobial regimens for pediatric bloodstream infections in Malawi (1998–2017). *Clin Infect Dis* 2019;69:61–68.
- Haigh K, Dube Q, Kasambara W, Feasey NA, Lester R. Cephalosporin resistance in Malawi. *Lancet Infect Dis* 2020;20:285–286.
- Musicha P, Feasey NA, Cain AK, Kallonen T, Chaguza C et al. Genomic landscape of extended-spectrum β -lactamase resistance in *Escherichia coli* from an urban African setting. *J Antimicrob Chemother* 2017;72:1602–1609.
- Clinical and Laboratory Standards Institute. M100 Performance Standards for Antimicrobial Susceptibility Testing. https://clsi.org/media/2663/m100ed29_sample.pdf (accessed 9 May 2020).
- nanoporetech. nanoporetech/flappie. *GitHub*. <https://github.com/nanoporetech/flappie> (accessed 3 March 2020).

30. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. Assembly of long, error-prone reads using repeat graphs. *Nat Biotechnol* 2019;37:540–546.
31. Wick RR, Holt KE. Benchmarking of long-read assemblers for prokaryote whole genome sequencing. *F1000Res* 2019;8:2138.
32. Tvedte ES, Gasser M, Sparklin BC, Michalski J, Zhao X et al. Comparison of long read sequencing technologies in resolving bacteria and fly genomes.
33. Vaser R, Sović I, Nagarajan N, Šikić M. Fast and accurate *de novo* genome assembly from long uncorrected reads. *Genome Res* 2017;27:737–746.
34. Website. Medaka: sequence correction provided by ONT research. <https://github.com/nanoporetech/medaka> [accessed 3 March 2020].
35. Inouye M, Dashnow H, Raven L-A, Schultz MB, Pope BJ et al. SRST2: rapid genomic surveillance for public health and hospital microbiology labs. *Genome Med.* 2014;6:90
36. Gonzalez-Escalona N, Allard MA, Brown EW, Sharma S, Hoffmann M. Nanopore sequencing for fast determination of plasmids, phages, virulence markers, and antimicrobial resistance genes in Shiga toxin-producing *Escherichia coli*.
37. Mohsin M, Raza S, Schaufler K, Roschanski N, Sarwar F et al. High prevalence of CTX-M-15-Type ESBL-producing *E. coli* from migratory avian species in Pakistan. *Front Microbiol.* 2017;8:2476
38. Shropshire WC, Aitken SL, Pifer R, Kim J, Bhatti MM et al. Concurrency of porin loss and modular amplification of β -lactamase encoding genes drives carbapenem resistance in a cohort of recurrent Enterobacterales bacteremia. *BioRxiv.*
39. Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 2018;34:3094–3100.
40. Naidenov B, Lim A, Willyerd K, Torres NJ, Johnson WL et al. Pan-genomic and polymorphic driven prediction of antibiotic resistance in *Elizabethkingia*. *Front Microbiol* 2019;10:1446.
41. Edgar RC. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 2004;5:113–119.
42. Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* 2019;35:4453–4455.
43. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 2012;19:455–477.
44. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S et al. Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics* 2015;31:3691–3693.
45. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc* 1995;57:289–300.
46. Website. <https://www.health.gov.mw/index.php/health-ethics-standards?download=57:mstg-full-cover2015> (accessed 16 May 2020).
47. Ministry of Health. Malawi National Tuberculosis Control Programme Manual. https://pdf.usaid.gov/pdf_docs/PA00KGDD.pdf (accessed 15 May 2020).
48. Feehily C, Karatzas KAG. Role of glutamate metabolism in bacterial responses towards acid and other stresses. *J Appl Microbiol* 2013;114:11–24.
49. Afset JE, Bruant G, Brousseau R, Harel J, Anderssen E et al. Identification of virulence genes linked with diarrhea due to atypical enteropathogenic *Escherichia coli* by DNA microarray analysis and PCR. *J Clin Microbiol* 2006;44:3703–3711.
50. Binns MM, Mayden J, Levine RP. Further characterization of complement resistance conferred on *Escherichia coli* by the plasmid genes *traT* of R100 and *iss* of ColVI-K94. *Infect Immun* 1982;35:654–659.
51. Binns MM, Davies DL, Hardy KG. Cloned fragments of the plasmid ColVI-K94 specifying virulence and serum resistance. *Nature* 1979;279:778–781.
52. Johnson TJ, Wannemuehler YM, Nolan LK. Evolution of the *iss* gene in *Escherichia coli*. *Appl Environ Microbiol* 2008;74:2360–2369.
53. Tarr PI, Bilge SS, Vary JC, Jelacic S, Habeeb RL et al. Iha: a novel *Escherichia coli* O157:H7 adherence-conferring molecule encoded on a recently acquired chromosomal island of conserved structure. *Infect Immun* 2000;68:1400–1407.
54. Guyer DM, Radulovic S, Jones F-E, Mobley HLT. Sat, the secreted autotransporter toxin of uropathogenic *Escherichia coli*, is a vacuolating cytotoxin for bladder and kidney epithelial cells. *Infect Immun* 2002;70:4539–4546.
55. Nataro JP, Seriwatana J, Fasano A, Maneval DR, Guers LD et al. Identification and cloning of a novel plasmid-encoded enterotoxin of enteroinvasive *Escherichia coli* and *Shigella* strains. *Infect Immun* 1995;63:4721–4728.
56. Mao B-H, Chang Y-F, Scaria J, Chang C-C, Chou L-W et al. Identification of *Escherichia coli* genes associated with urinary tract infections. *J Clin Microbiol* 2012;50:449–456.
57. Tyler AD, Mataseje L, Urfano CJ, Schmidt L, Antonation KS et al. Evaluation of Oxford Nanopore's Minion sequencing device for microbial whole genome sequencing applications. *Sci Rep.* 2018;8
58. Bowden R, Davies RW, Heger A, Pagnamenta AT, de Cesare M et al. Sequencing of human genomes with nanopore technology. *Nat Commun* 2019;10:10.

Five reasons to publish your next article with a Microbiology Society journal

1. The Microbiology Society is a not-for-profit organization.
2. We offer fast and rigorous peer review – average time to first decision is 4–6 weeks.
3. Our journals have a global readership with subscriptions held in research institutions around the world.
4. 80% of our authors rate our submission process as 'excellent' or 'very good'.
5. Your article will be published on an interactive journal platform with advanced metrics.

Find out more and submit your article at microbiologyresearch.org.