

RESEARCH

Open Access



Unraveling virulence determinants in extended-spectrum beta-lactamase-producing *Escherichia coli* from East Africa using whole-genome sequencing

Ivan Sserwadda^{1,2,3†}, Benson R. Kidenya^{4†}, Stephen Kanyerezi^{1,3}, Inyasi Lawrence Akaro⁵, Baraka Mkinze⁵, Stephen E. Mshana⁶, Suhaila O. Hashim², Everlyne Isoe², Jeremiah Seni⁶, Moses L. Joloba¹ and Gerald Mboowa^{1,3,7*}

Abstract

Escherichia coli significantly causes nosocomial infections and rampant spread of antimicrobial resistance (AMR). There is limited data on genomic characterization of extended-spectrum β -lactamase (ESBL)-producing *E. coli* from African clinical settings. This hospital-based longitudinal study unraveled the genetic resistance elements in ESBL *E. coli* isolates from Uganda and Tanzania using whole-genome sequencing (WGS). A total of 142 ESBL multi-drug resistant *E. coli* bacterial isolates from both Tanzania and Uganda were sequenced and out of these, 36/57 (63.1%) and 67/85 (78.8%) originated from Uganda and Tanzania respectively. Mutations in *RarD*, *yaaA* and *ybgI* conferring resistances to chloramphenicol, peroxidase and quinolones were observed from Ugandan and Tanzanian isolates. We reported very high frequencies for *bla*_{CTX-M-15} with 11/18 (61.1%), and *bla*_{CTX-M-27} with 12/23 (52.1%), *bla*_{TEM-1B} with 13/23 (56.5%) of isolates originating from Uganda and Tanzania respectively all conferring resistance to Beta-lactam-penicillin inhibitors. We observed chloramphenicol resistance-conferring gene *mdfA* in 21/23 (91.3%) of Tanzanian isolates. Extraintestinal *E. coli* sequence type (ST) 131 accounted for 5/59 (8.4%) of Tanzanian isolates while enterotoxigenic *E. coli* ST656 was reported in 9/34 (26.4%) of Ugandan isolates. Virulence factors originating from *Shigella dysenteriae* Sd197 (*gspC*, *gspD*, *gspE*, *gspF*, *gspG*, *gspH*, *gspI*), *Yersinia pestis* CO92 (*irp1*, *ybtU*, *ybtX*, *iucA*), *Salmonella enterica* subsp. *enterica* serovar Typhimurium str. LT2 (*csgF* and *csgG*), and *Pseudomonas aeruginosa* PAO1 (*fliA*, *fliG*, *fliM*) were identified in these isolates. Overall, this study highlights a concerning prevalence and diversity of AMR-conferring elements shaping the genomic structure of multi-drug resistant *E. coli* in clinical settings in East Africa. It underscores the urgent need to strengthen infection-prevention controls and advocate for the routine use of WGS in national AMR surveillance and monitoring programs.

Availability of WGS analysis pipeline: the rMAP source codes, installation, and implementation manual can free be accessed via <https://github.com/GunzIvan28/rMAP>.

[†]Ivan Sserwadda and Benson R. Kidenya contributed equally

*Correspondence:
Gerald Mboowa
gmboowa@gmail.com

Full list of author information is available at the end of the article



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Keywords Extended-spectrum β -lactamase, East Africa, Whole-genome sequencing, Antimicrobial resistance, Virulence factors

Background

The antimicrobial resistance (AMR) phenomenon has spread rapidly over the course of the past decades to establish itself as a major global public health threat [1] in spite of the strides made by modern medicine to apply the use of antibiotics to ensure safe surgical procedures and improve the quality of medical care that have greatly reduced morbidity and mortality in public health [2, 3]. The bacterial infections that have been able to exhibit AMR have become fatal and given birth to a possible post-antibiotic era which initially was thought to be an apocalyptic fantasy before the 21st century [2, 4]. It goes without saying that the role played by humans in exacerbating the rate at which these infectious agents have developed resistance to antibiotics has been at the forefront of accelerating this phenomenon mainly through the rampant and inappropriate use of antibiotics [4, 5]. A previous study predicted an alarming 10 million deaths per annum with \$100 trillion dollars' worth of efforts trying to combat AMR by 2050 if this is not tackled [6]. Some regions in the world especially the African continent characterized by high infectious disease burdens and limited healthcare infrastructure have the least accurate and reliable statistical data on the epidemiology and impact of AMR on the public health sectors [7–9]. The rapid recent evolution of genomics-based technologies applied in the diagnosis and surveillance of the epidemiology of drug-resistant bacteria has led to the generation of large amounts of genomic data that have given deeper insights into the nature and changes of AMR determinants using modern bioinformatics analysis pipelines [10]. Application of next-generation sequencing technologies (NGS) alongside conventional microbiology procedures and antimicrobial susceptibility testing (AST) may be the key to understanding AMR [11], accelerating knowledge generation, and deploying interventions tailored towards optimization of antimicrobial use in public health [12].

Therefore, this study explored and sought to understand the epidemiology, and factors driving AMR in hospital settings in Uganda and Tanzania through the use of whole-genome sequencing (WGS) data combined with the socio-demographic metadata provided by the mother study.

Materials and methods

Study design and settings

This study utilized a laboratory-based and longitudinal study design approach. The laboratory-based design was

used to undertake WGS to determine the AMR elements from the bacterial isolates.

Study sites

This study was carried out at the orthopedic units of Mulago National Referral Hospital, Kampala, Uganda with geographical coordinates (0°20'16.0"N, 32°34'32.0"E) and Bugando Medical Centre (BMC), Mwanza, Tanzania with geographical coordinates (2°31'41.0"S, 32°54'27.0"E).

Study population

The study population constituted of whole-genome sequence data obtained from a total of 142 multi-drug resistant *E. coli* bacterial isolates provided by the mother study titled, "Understanding Transmission Dynamics and Acquisition of Antimicrobial Resistance at Referral Hospitals and Community Settings in East Africa using Conventional Microbiology and Whole-Genome Sequencing". The multi-drug isolates were collected from both study sites in Uganda (n=57) and Tanzania (n=85) from patients, the immediate non-medical caretakers of these patients, the immediate health workers attending to these patients and the patients' environment as previously discussed in detail by the recent publications from Uganda [1] and Tanzania [13].

Data collection and analysis tools

The WGS data used in this study was provided by a bigger mother study titled, "Understanding Transmission Dynamics and Acquisition of Antimicrobial Resistance at Referral Hospitals and Community settings in East Africa using Conventional Microbiology and Whole-Genome Sequencing (Grant number GCA/AMR/rnd2/058)". The bacterial isolates were shipped and sequenced by the Earlham Institute, Norwich, located in the United Kingdom following the Low Input, Transposase Enabled (LITE) Illumina protocol using the Illumina NovaSeq 6000 System.

The analysis of WGS data was done using our previously published Linux command line-based bioinformatics workflow called "rMAP", the Rapid Microbial Analysis pipeline [10]. Briefly, the whole-genome raw sequences together with the GenBank *Escherichia coli* str. K-12 substr. MG1655 with Accession NC_000913 reference was fed into the rMAP pipeline. Sequences in the format fastq.gz were used as the input for the pipeline. All sequences were inspected for quality in the rMAP pipeline [10] before any subsequent processes using the embedded FastQC [14] to generate individual sample reports and MultiQC [15] for aggregating all the multiple

reports into one report. Adapters were trimmed off the sequences using Trimmomatic [16] with the selected parameters including minimum length and phred score set to 200 and 32 respectively.

The trimmed reads were loaded into the Shovill [17] *de-novo* pipeline using the Skesa as the assembler of choice. *K*-mer sizes of 31, 55, 79, 103 and 127 were used to determine the optimum genome assembly. Pilon was used for checking assembly errors, correcting ambiguous gaps, insertions, deletions and finally polishing the genomes [18]. Genome annotations were performed using the Prokka tool [19].

Single nucleotide polymorphism (SNP) variant calling was performed using SAMtools, Burrows-Wheeler Aligner (BWA), SAMclip, Freebayes and SnpEff [20–22]. The trimmed reads were aligned against an indexed reference in fasta format (GenBank reference *Escherichia coli* strain K-12 sub strain MG1655 Accession: NC_000913.3) using Burrows-Wheeler aligner [21] to produce Sequence Alignment Map (SAM) files. Soft and hard clipped alignments were removed from the SAM files using SAMclip (<https://github.com/tseemann/samclip>). SAMtools [20] then sorted, marked duplicates and indexed the resultant Binary Alignment Map (BAM) files. Freebayes [23] was used to call variants using Bayesian models to produce variant call format (VCF) files containing SNP information which was filtered using BCFtools (<https://github.com/samtools/bcftools>) and normalized of biallelic regions using Vt [24]. The filtered VCF files were annotated using SnpEff [22]. Missense variants that were associated with resistance were identified from the VCF files according to their respective sites. Only true SNPs were considered for the downstream analysis; insertions, deletions, and complex SNPs were filtered out from the resistance-associated SNPs.

Phylogenetic inference by maximum likelihood was performed using MAFFT, IQtree, Vcf2phylip, and BMGE [25–28]. The rMAP pipeline collated all the individual VCF files into a single VCF containing all the samples and their SNPs before being transposed by vcf2phylip [26] into a multi-alignment fasta file. MAFFT software package was used to perform multiple sequence alignment [27]; removal of ambiguously aligned reads as well as extraction of informative sites was performed to infer phylogeny using BMGE [25]. IQtree [28] was then used to test various substitution models and construct trees from the alignments using the maximum-likelihood method in 1,000 bootstraps. The resulting trees were visualized in the form of rectangular phylograms.

Mass screening for AMR genes against CARD [29], ARG-ANNOT [30], NCBI, ResFinder, and MEGARES [31] databases was performed for each of the study isolates using the Abricate tool (<https://github.com/tseemann/abricate>). For consistency purposes, we compared

results from the two most commonly used well-annotated AMR databases across the *E. coli* isolates (CARD and ResFinder) from both study sites. From our findings, we found more AMR genes conforming to >90% cut-off for both coverage and identity being detected from the ResFinder database which were then presented in form frequencies and heatmaps in the results section.

Multi-locus sequence typing was performed using MLST (<https://github.com/tseemann/mlst>) from the *E. coli* assembled contigs. The *E. coli* isolates were determined against a set of seven (7) *E. coli* housekeeping genes (*adk4*, *fum26*, *gyrB2*, *icd25*, *mdh5*, *purA5*, and *recA19*).

Virulence factors were determined against the virulence factor database (VFDB) [32] using the Abricate tool (<https://github.com/tseemann/abricate>) to identify elements that conformed to >90% cut-off for both coverage and identity.

Results

Pre-sequencing quality control

A total of 142 ESBL multi-drug resistant *E. coli* bacterial isolates from both Tanzania and Uganda were sequenced. Only 36/57 (63.1%) of the ESBL-producing *E. coli* from Uganda and 67/85 (78.8%) of the ESBL-producing *E. coli* isolates from Tanzania passed the initial pre-sequencing Transposase Enabled (LITE) Illumina protocol and were selected for the downstream bioinformatics analysis as shown in Fig. 1.

Raw sequence quality assessment, adapter, and poor read sequence trimming

The average read length of the sequences was 150 bp with an average Q-score of 32 as discussed in the methods section.

De-novo genome assembly

Considering the main genome assembly parameters, the samples from the two study sites had an average of 132 bp (base pairs) for the mean read length and 525 bp for the contigs. An estimated average genome length of 5,138,608 bp, average GC-content of 51%, and average sequencing depth of 9.34X were reported for the isolates. A summary of the average genome assembly statistics across all analyzed samples is shown in Table 1.

Single nucleotide polymorphism (SNP) variant calling

The annotated VCF files were interrogated for SNPs that were “missense” meaning those that altered the function of the protein and had a keyword “resistance”. Generally, the most clinically significant predominant resistance-associated SNPs and respective genes affected included: peroxide stress resistance protein (*yaaA*), chloramphenicol resistance permease (*RarD*), and quinolone resistance

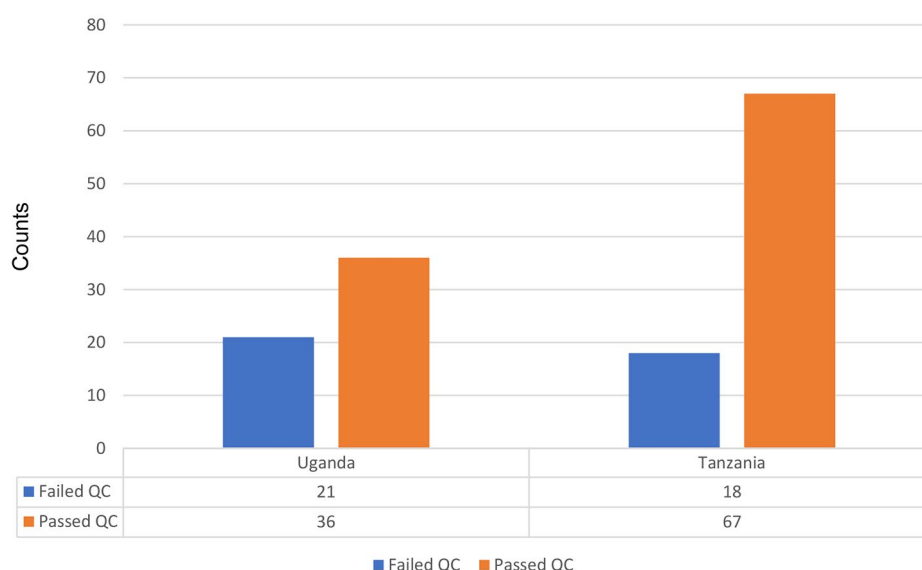


Fig. 1 Proportions of the ESBL-producing *E. coli* from the study sites that conformed to initial pre-sequencing (LITE) Illumina protocol

Table 1 *De novo* genome assembly metrics and statistics for the study isolates

Metric	Uganda	Tanzania
Trimmed read counts	359,277	405,411
Mean Read Length (base pairs, bp)	133	131
Number of Contigs (> 200 bp)	612	437
Approximates Genome Length (bp)	5,034,750	5,242,466
Largest Contig (bp)	240,554	340,237
N50	80,279	116,894
GC- Content (%)	51.051	51.046
Depth (X)	8.85	9.822

Table 2 Frequency of resistance-associated SNPs identified from Ugandan and Tanzanian isolates

Resistance gene	Functional annotation and conferred resistance	Uganda (n = 36)	Tanzania (n = 67)
<i>RarD</i>	Chloramphenicol resistance permease	14/36 (38.8%)	23/67 (34.3%)
<i>yaaA</i>	Peroxide stress resistance protein	11/36 (30.5%)	33/67 (49.2%)
<i>ybgI</i>	Quinolone resistance protein and radiation resistance protein	25/36 (69.4%)	30/67 (44.7%)

protein (*ybgI*). A detailed depiction of the distribution of these SNPs between the two study sites is shown in Table 2.

Antimicrobial resistance gene determinants from Uganda and Tanzania

From the Ugandan site, 18/36 (50%) isolates were found to have AMR genes conforming to the set >90% cut-off for both coverage and identity in the ResFinder database. The most predominant genes identified among the isolates comprised of *bla*_{CTX-M-15} with 11/18(61.1%)

which confers resistance to fluoroquinolones, and third-generation cephalosporins, *mdfA* 14/18(77.7%) which confers resistance to chloramphenicol through its efflux pump action, *tet(B)* 9/18 (50%) conferring resistance to tetracyclines and *sul1* 8/18(44.4%) conferring resistance to sulphonamides. The heatmap in Fig. 2 shows the distribution of the respective genes across the isolates. The green color indicates the presence of a gene while the grey shows the absence of that gene per isolate.

From the Tanzanian site, 23/67 (34.3%) isolates were found to have AMR genes conforming to the set >90% cut-off for both coverage and identity in the ResFinder database. The most predominant genes identified among the isolates comprised of *bla*_{CTX-M-27} with 12/23 (52.1%) conferring resistance to fluoroquinolones and third-generation cephalosporins, and *bla*_{TEM-1B} with 13/23 (56.5%) which confers resistance to third-generation cephalosporins, *mdfA* (A) 21/23 (91.3%) which confers resistance to chloramphenicol through its efflux pump action, *tet(A)* 12/23 (52.1%) conferring resistance to tetracyclines and *sul2* 15/23 (65.2%) conferring resistance to sulphonamides. The heatmap in Fig. 3 shows the distribution of the respective genes across the isolates. The green color indicates the presence of a gene while the grey shows the absence of that gene per isolate.

Multi-locus sequence typing (MLST)

A total of 34/36 (94.4%) isolates for which MLST was performed conformed to the set >90% cut-off for both coverage and identity for the sequence types (STs). The most predominant *E. coli* from Uganda comprised of the following: ST656 accounting for 9/34 (26.4%) within the isolates, ST206, ST448, ST1193, and ST1284 each

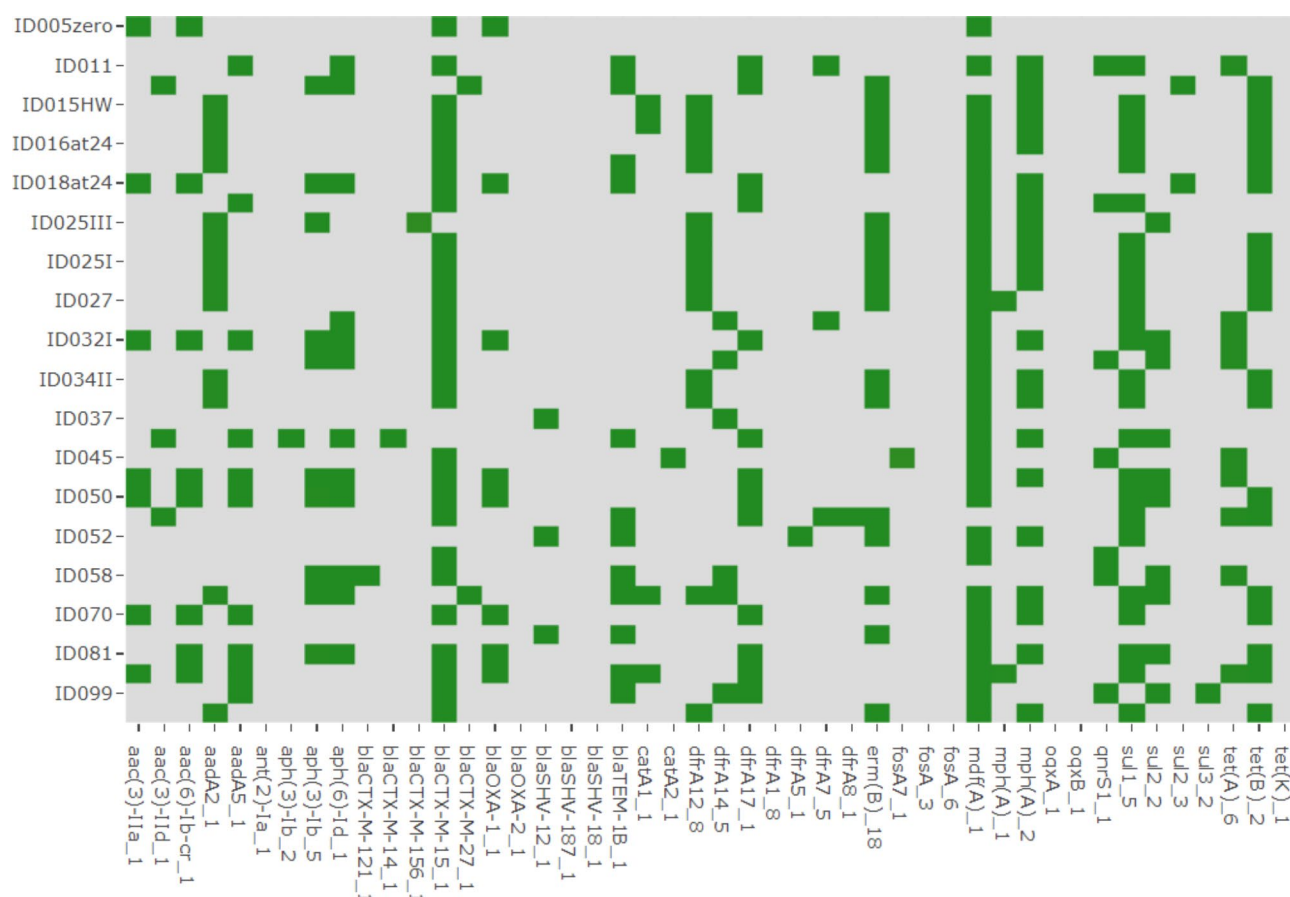


Fig. 2 Heat map showing Ugandan *E. coli* predominant genes from ResFinder database

accounting for 2/34 (5.8%) within the isolates. The majority of the other STs were; ST131, ST10, ST6438, ST998, and ST38 among others each accounted for 1/34 (2.9%) within the isolates.

In comparison with the Tanzanian site, 59/67 (88%) isolates conformed to the set >90% cut-off for both coverage and identity for the sequence types (STs). The most predominant *E. coli* STs from Tanzania comprised of the following: ST131 accounting for 5/59 (8.4%) within the isolates, ST10, ST2852, and ST167 each accounting for 2/59 (3.3%) within the isolates. Majority of the other STs were: ST3580, ST695, ST542, ST48, and ST38 among others accounted for 1/59 (1.6%) within the isolates. Fig. 4 below shows how the isolates between the two countries cluster in terms of AMR genes and STs.

Virulence factor detection(VF)

In addition to their AMR genes, the *E. coli* species from the two study sites were found to possess arsenals of virulence factors that are inter-species in nature that is, they originated from other different bacterial species, a clear depiction of horizontal transfer of genetic material between bacteria. The isolates from the two countries had most of their virulence factors identical to those found in

Shigella dysenteriae Sd197, *Yersinia pestis* CO92, *Salmonella enterica* subsp. enterica serovar Typhimurium str. LT2, *Escherichia coli* O157:H7 str. EDL933, and *Pseudomonas aeruginosa* PAO1. A comprehensive table depicting the distribution of the different virulence factors of the isolates from the two sites is shown in Table 3.

Discussion

Control of multi-drug resistant infections is fundamental in reducing the disease burden and costs incurred while treating these pathogens in tandem with the Global Action Plan set by WHO [33] to combat AMR. This study comes in at the right point in time where the scale at which global public health is threatened by the increasing infection rates. In this study, the aim was to explore the genetic determinants that confer AMR from isolates obtained at Mulago National Referral Hospital, Bugando Medical Centre, and their environmental settings. The findings from this study are provocative and inform the dire need to strengthen the existing infection-prevention controls (IPC) together with surveillance and monitoring systems.

This study was predominantly comprised of ESBL organisms; with 36/57 (63.1%) originating from Uganda

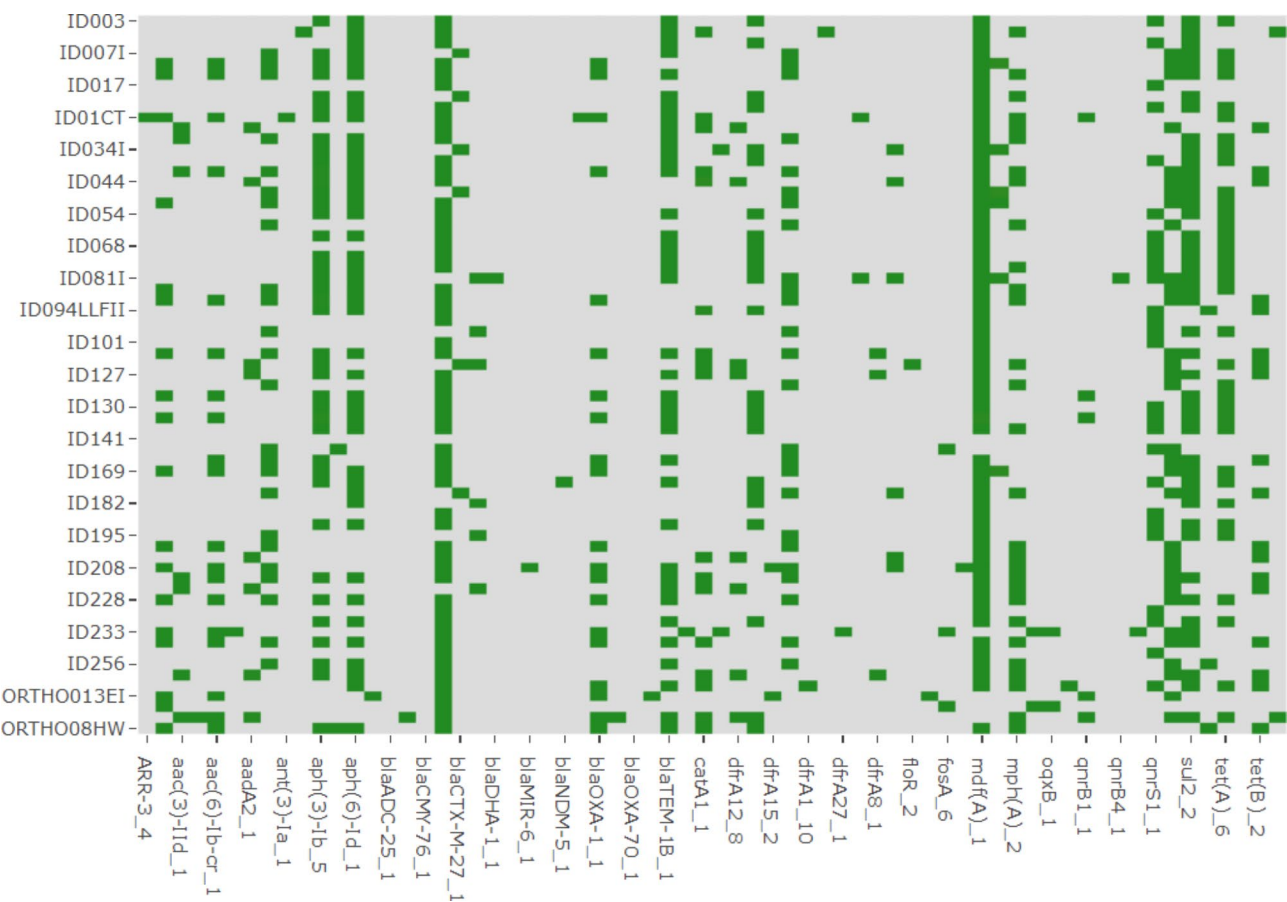


Fig. 3 Heat map showing Tanzanian *E. coli* predominant genes from ResFinder database

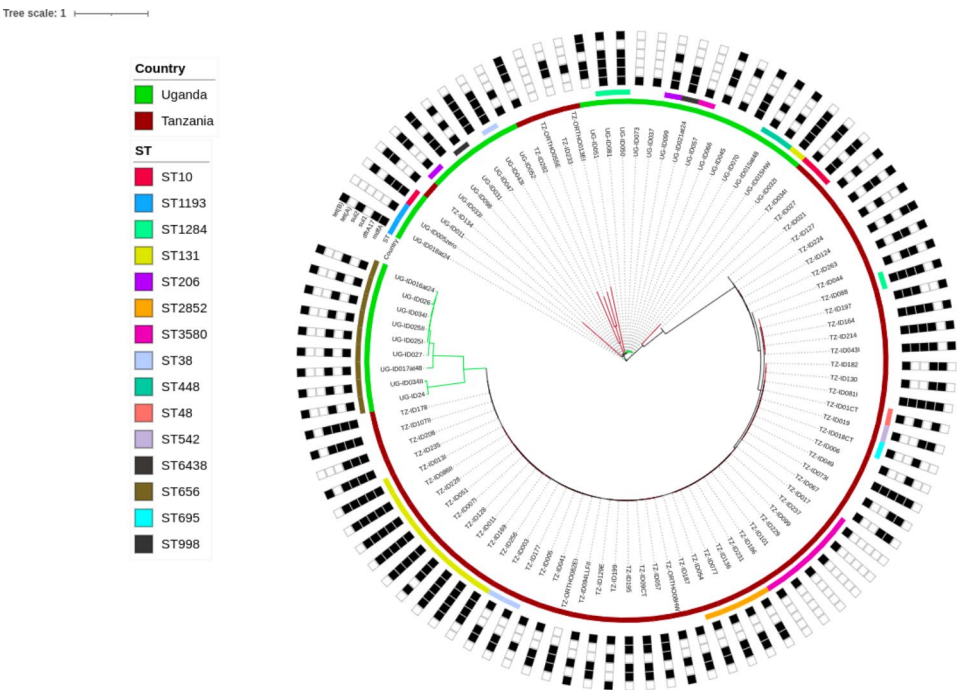


Fig. 4 Phylogenetic tree showing the distribution of AMR genes and STs across the two study sites

Table 3 Distribution of virulence factors among the *E. coli* isolates from Uganda and Tanzania

Virulence Factors (VF)	Physio-biological roles	Isolates	
		Uganda	Tanzania
<i>gspD</i> , <i>gspE</i> , <i>gspF</i> , <i>gspG</i> , and <i>gspH</i>	General secretion pathway proteins for <i>Shigella dysenteriae</i> Sd197	✓	✓
<i>irp1</i> , <i>ybtU</i> , <i>ybtX</i> and <i>iucA</i>	A yersiniabactin biosynthetic protein for <i>Yersinia pestis</i> CO92	✓	✓
<i>csgF</i> and <i>csgG</i>	Curli production assembly/transport proteins for <i>Salmonella enterica</i> subsp. <i>enterica</i> serovar Typhimurium str. LT2	✓	✓
<i>gspC</i> , <i>gspH</i> , and <i>gspI</i>	General secretion pathway proteins C for <i>Shigella dysenteriae</i> Sd197	✓	✓
<i>espL1</i> , <i>espL4</i> , <i>espR3</i> , <i>espR4</i> , <i>espX1</i> , <i>espX4</i> , <i>espY1</i> , and <i>espY3</i>	Type III secretion system effectors for <i>Escherichia coli</i> O157:H7 str. EDL933	✓	✓
<i>astA</i>	A heat-stable enterotoxin 1 for <i>Escherichia coli</i> O44:H18 O42	✓	–
<i>eltA</i> and <i>eltB</i>	Heat-labile enterotoxins from human for <i>Escherichia coli</i>	✓	–
<i>afaB-I</i> and <i>afaC-I</i>	Chaperone mannose-resistant adhesin proteins for <i>Escherichia coli</i> O25b:H4 str. FV9863	✓	–
<i>fepA</i> , <i>fepB</i> , <i>fepC</i> , <i>fepD</i> , and <i>fepG</i>	Ferrienterobactin outer membrane transporter permease Enterobactin for <i>Escherichia coli</i> CFT073	–	✓
<i>ompA</i>	Outer membrane protein for <i>Escherichia coli</i> O18:K1:H7 str. RS218	–	✓
<i>hlyA</i> , <i>hlyB</i> , <i>hlyC</i> , and <i>hlyD</i>	Hemolysin A for <i>Escherichia coli</i> CFT073	–	✓

and 67/85 (78.8%) originating from Tanzania. Previous findings from studies in Uganda reported ESBL *E. coli* prevalence rates of 5.3% conducted between 2006 and 2007 [34], 62% carried out in 2014 [35] at Mulago National Referral Hospital, and between 2015 and 2016 at Kasese Regional Referral Hospital at 62% [36]. Related meta-analysis and systematic review studies from East Africa carried out in hospitals and surrounding communities reported a similar predominance of ESBL-producing *Escherichia coli* and *Klebsiella pneumoniae* [37–39]. The Enterobacteriaceae family has been reported to shape the nosocomial pathogen eco-system because of the plasticity of their genome and their ability to perform inter-species and intra-species incorporation and transfer of drug resistance mediating determinants like plasmids, transposons, insertion sequences, and virulence factors via horizontal gene transfer [40–42].

Detection of SNP-associated mutations in the genes; *RarD*, *yaaA*, and *ybgI* conferring resistances to chloramphenicol, peroxidase, and quinolones from the isolates

from both Uganda and Tanzania further depicts how these organisms evolve resistance towards some of the most commonly used antibiotics and antiseptic used for the day-to-day management of clinical cases. A related study highlighted the roles played by these SNPs in the evolution of antimicrobial resistance and in shaping the genome of these organisms [43].

Our results reported very high frequencies for *bla*_{CTX-M-15} accounting for 11/18(61.1%), and *bla*_{CTX-M-27} with 12/23 (52.1%), *bla*_{TEM-1B} with 13/23 (56.5%) of isolates originating from Uganda and Tanzania respectively. These genes are responsible for conferring resistance to penicillin, fluoroquinolones, and third-generation cephalosporins (ceftazidime and cefotaxime) which are part of routinely prescribed antibiotics used in the treatment of medical cases within the two sites similar to other previous related studies that reported phenotypic AMR profiles of the same organisms [44]. Tanzania had relatively higher tetracycline resistance gene *tet(A)* with 12/23 (52.1%) as compared to the Ugandan isolates with *tet(A)* with 7/18 (38.8%) prevalence which are in agreement with similar studies performed across six Tanzanian hospitals [45]. We also reported relatively a high prevalence of sulphonamide-resistance conferring genes *sul1* 8/18(44.4%) and *sul2* 15/23 (65.2%) from Uganda and Tanzania respectively. Chloramphenicol resistance gene *mdfA(A)* with 21/23 (91.3%) from Tanzanian isolates and trimethoprim resistance-conferring gene *dfrA17* with 8/23 (34.7%) in Ugandan isolates were also observed within the two cohorts. The authors propose that the high resistance observed in the majority of over-the-counter antibiotics can likely be explained by their affordability in the two countries. In contrast, expensive drugs such as piperacillin-tazobactam, amikacin, and carbapenems are less accessible in most drug shops within the two study sites. Consequently, these expensive drugs are less likely to be misused, leading to lower detected resistance levels. These findings sound a very big alarm about the potential dangers such pathogens can cause to the general public health and call for the need to scale up the microbiology laboratory capacity as a way of guiding antimicrobial agent prescription. Data from well-established laboratory facilities will shape and strengthen AMR surveillance, IPC protocols within community-based settings and healthcare facilities, and regulation of drug prescriptions from drug outlets and pharmacies.

The largest portion of sequence types isolated from Tanzania belonged to the ST131 accounting for 5/59 (8.4%) of the total isolates while Uganda was represented with 1/34 (2.9%) for ST131 of the total isolates. These sequence types associated with extra-intestinal infections have been reported to be rapidly spreading as high-risk clones in Europe and worldwide due to increased AMR [46–49]. Some other sequence types like ST206 have

been reported to be associated with colistin-resistance conferring isolates from a study in China [50]. These findings reiterate the dangers these organisms impose on the public health system and call for immediate interventions [51].

The presence of virulence factors like *Shigella dysenteriae* Sd197 (*gspD*, *gspE*, *gspF*, *gspG*, and *gspF*), *Yersinia pestis* CO92 (*irp1*, *ybtU*, *ybtX*, and *iucA*), *Salmonella enterica* subsp. *enterica* serovar Typhimurium str. LT2 (*csgF* and *csgG*), *Shigella dysenteriae* Sd197 (*gspC*, *gspD*, *gspE*, *gspF*, *gspG*, *gspH*, and *gspI*), and *Pseudomonas aeruginosa* PAO1 (*flhA*, *fliG*, and *fliM*) in isolates from both Tanzania and Uganda reported by this study depict a classic case of inter-species genetic-determinant element transfer. Multiple studies have reported how the horizontal gene transfer, a process through which genetic information can be acquired from the environment to a bacterium or from one bacterium to another through other mechanisms other than chromosomal inheritance consequently shaping pathogen virulence evolution [52–56]. This study provides strong evidence regarding the acquisition of a set of rather queer virulence factors like *Yersinia pestis* CO92 among the *Escherichia coli* isolates originating from a somewhat deadly plague-causing bacteria similar to what has been reported by a study in India [57].

The ingenious advent of bioinformatics platforms like rMAP [10] used to profile all these virulence elements within the isolates in one go provides a comprehensive way of analyzing WGS data because of its easy installation, usage, and applicability, especially in low-income settings where high-performance computing infrastructure is limited. In our opinion, it also bridges and fills the missing link between the rapidly embraced field of WGS and conventional microbiology while providing high-resolution, and shorter result-generating turnaround times for the genomes of MDR pathogens. It is on these grounds that we recommend this tool to be adopted as a continuous monitoring and surveillance software for monitoring the antimicrobial resistance gene trends, plasmids, virulence factors, and MLSTs within community and healthcare settings for Uganda, Tanzania, and Africa as a whole.

Conclusion

This study shows a notable abundance and diversity of the AMR-conferring elements shaping the genomic structure and survival of *E. coli* isolates from the hospital and clinical settings in Uganda and Tanzania. The findings from this study are provocative and inform the dire need to strengthen the existing infection-prevention controls (IPC) and adoption of WGS alongside conventional microbiology as a way of building genomics surveillance and monitoring capacity in East Africa.

Limitations

Due to financial limitations, our study relied exclusively on WGS short-read data for deductions. Introducing long-read sequencing such as Nanopore sequencing requires a low quantity of input DNA and is suitable for low-concentration samples and in conjunction with short-read might have enriched our understanding of the AMR-conferring elements elucidated in our research. Hybrid long-short-read sequencing approaches offer a powerful tool in microbial genomic research. By combining the strengths of both long-read and short-read sequencing, hybrid approaches offer improved accuracy, read length, and cost-effectiveness.

Abbreviations

AMR	antimicrobial resistance
CUHAS	Catholic University of Health and Allied Sciences
ESBL	extended-spectrum beta-lactamase
HGT	horizontal gene transfer
IPC	infection and prevention control
MDR	multi-drug resistance
PDR	pan-drug resistance
rMAP	Rapid Microbial Analysis Pipeline
STs	sequence types

Acknowledgements

Ivan Sserwadda is a fellow of the Makerere University/Uganda Virus Research Institute (UVRI) Centre of Excellence for Infection & Immunity Research and Training (MUII) program. He is also supported by MUII. MUII is supported through the DELTAS Africa Initiative (Grant no. 107743). The DELTAS Africa Initiative is an independent funding scheme of the African Academy of Sciences (AAS), Alliance for Accelerating Excellence in Science in Africa (AESA) and supported by the New Partnership for Africa's Development Planning and Coordinating Agency (NEPAD Agency) with funding from the Wellcome Trust (Grant no. 107743) and the UK Government. We are equally grateful for the support of the NIH Common Fund, through the OD/Office of Strategic Coordination (OSC) and the Fogarty International Center (FIC), NIH award number U2RTW010672. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the supporting offices.

Author contributions

IS, GM, JS, and BRK made substantial contributions to the conception, designing, coordination, and execution of the study; IS, SK, and GM participated in the analysis and interpretation of the data; IS, GM, JS, ILA, BM, SOH, EI, MLJ, SEM, and BRK wrote the initial draft of the manuscript, which was critically revised by all authors. All the authors reviewed and approved the final manuscript.

Funding

This study was part of the research project titled, "Understanding transmission dynamics and acquisition of antimicrobial resistance at referral hospitals and community settings in East Africa", supported through the Grand Challenges (GC) Africa programme (Grant number GCA/AMR/rnd2/058). The GC Africa is a programme of the African Academy of Sciences (AAS) implemented through the Alliance for Accelerating Excellence in Science in Africa (AESA) platform, an initiative of the AAS and the African Union Development Agency (AUDA-NEPAD). The GC Africa is supported by the Bill and Melinda Gates Foundation (BMGF) and the African Academy of Sciences and partners. It was executed collaboratively between Makerere University College of Health Sciences, Kampala, Uganda, and the Catholic University of Health and Allied Sciences, Mwanza, Tanzania. The funders had no roles in the design, execution, data management, manuscript writing, or manuscript submission. This work was also in part supported by the African Association for Research and Control of Antimicrobial Resistance (AAAMR) 2020 Young Anglophone Investigator Award to G.M and the Public Health Alliance for Genomic Epidemiology

(PHA4GE) Grant (INV-038071) to support pathogen genomics data standards and bioinformatics interoperability awarded to I.S.

Data Availability

All data generated or analyzed during this study are included in this published article. All source code for the rMAP pipeline, installation instructions, and implementation can be accessed via GitHub (<https://github.com/Gunzlvan28/rMAP>). The source code is available on GitHub under the GPL3 license. Questions, bugs, or any other issues can be filed as GitHub issues. Although rMAP itself is published and distributed under a GPL3 license, some of its dependencies bundled within the rMAP volume are published under different license models.

The genomic raw reads files from this study are publicly available at the Sequence Read Archive (SRA) of the National Center for Biotechnology Information (NCBI) under the study BioProject ID: PRJNA840888.

Declarations

Ethics approval and consent to participate

The study was performed in accordance with the Declaration of Helsinki and was approved by the School of Biomedical Sciences-Research and Ethics Committee, College of Health Sciences, Makerere University (approval number: SBS-HDREC-650), Mulago National Referral Hospital-Research and Ethics Committee (approval number: MHREC-1702), and Uganda National Council for Science and Technology (approval number: HS411ES). It was also approved by the Joint CUHAS/BMC Research and Ethics Committee (CREC/409/2019) and the National Health Research Ethics Review Committee of the National Institute for Medical Research (NIMR/HQ/R.8a/Vol.IX/3322) in Tanzania. All information obtained was coded and kept confidential. Written informed consent was obtained from all the participants and from the legal guardians of minor and illiterate participants. For participants aged between 12 and 17 years, their assent was also obtained. Illiterate participants added their fingerprints to the consent form along with an impartial witness who had also signed and was present during the informed consent process.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Department of Immunology and Molecular Biology, School of Biomedical Sciences, College of Health Sciences, Makerere University, P.O. Box 7072, Kampala, Uganda

²Department of Biochemistry and Bioinformatics, School of Pure and Applied Sciences, Pwani University, P.O. Box 195-80108, Kilifi, Kenya

³The African Center of Excellence in Bioinformatics and Data-Intensive Sciences, Infectious Diseases Institute, College of Health Sciences, Makerere University, P.O. Box 22418, Kampala, Uganda

⁴Department of Biochemistry and Molecular Biology, Weill Bugando School of Medicine, Catholic University of Health and Allied Sciences, Mwanza, Tanzania

⁵Department of Surgery, Weill Bugando School of Medicine, Catholic University of Health and Allied Sciences, Mwanza, Tanzania

⁶Department of Microbiology and Immunology, Weill Bugando School of Medicine, Catholic University of Health and Allied Sciences, Mwanza, Tanzania

⁷Africa Centres for Disease Control and Prevention, African Union Commission, Roosevelt Street, P.O. Box 3243, Addis Ababa W21 K19, Ethiopia

References

1. Mboowa G, Sserwadda I, Bulafu D, Chaplain D, Wewedru I, Seni J, et al. Transmission Dynamics of Antimicrobial Resistance at a National Referral Hospital in Uganda. *Am J Trop Med Hyg*. 2021;105(2):498–506.
2. Rossolini GM, Arena F, Pecile P, Pollini SJCoip. Update on the Antibiotic Resistance Crisis. 2014;18:56–60.
3. Nathan C, Cars OJNEJoM. Antibiotic resistance—problems, progress, and prospects. 2014;371(19):1761–3.
4. Pärnänen K, Karkman A, Tamminen M, Lyra C, Hultman J, Paulin L et al. Evaluating the mobility potential of antibiotic resistance genes in environmental resistomes without metagenomics. 2016;6:35790.
5. McEwen SA, Collignon PJ, Aarestrup FM, Schwarz S, Shen J, Cavaco L. Antimicrobial Resistance: One Health Perspective. 2018;6(2):6. <https://doi.org/10.1128/microbiolspec.ARBA-0009-2017>. 2.10.
6. Jasovský D, Littmann J, Zorzet A, Cars O. Antimicrobial resistance—a threat to the world's sustainable development. 2016;121(3):159 – 64.
7. Ampaire L, Muhindo A, Orikiriza P, Mwanga-Amumpaire J, Bebell L, Boum YJAJolm. *Rev Antimicrob Resist East Afr*. 2016;5(1):1–6.
8. Arif M, Othman AA. Chapter 11 - Phenotype screenings of drugs for combination therapy against multidrug resistance. 2020:207 – 20. <https://doi.org/10.1016/B978-0-12-820576-1.00011-4>.
9. O'Neill J. Antimicrobial Resistance: tackling a crisis for the health and wealth of nations. *Rev Antimicrob Resist*. 2014:1–20.
10. Sserwadda I, Mboowa G. rMAP: the Rapid Microbial Analysis Pipeline for ESKAPE bacterial group whole-genome sequence data. *Microb Genomics*. 2021;7(6):000583. <https://doi.org/10.1099/mgen.0.000583>. PubMed PMID: 34110280.
11. Sserwadda I, Lukenge M, Mwambi B, Mboowa G, Walusimbi A, Segujja F. Microbial contaminants isolated from items and work surfaces in the post-operative ward at Kawolo general hospital, Uganda. *BMC Infect Dis*. 2018;18(1):68. <https://doi.org/10.1186/s12879-018-2980-5>. PubMed PMID: 29409447.
12. Mshana SE, Sindato C, Matee MI, Mboera LEG. Antimicrobial Use and Resistance in Agriculture and Food Production Systems in Africa: a systematic review. *Antibiotics*. 2021;10(8):976. <https://doi.org/10.3390/antibiotics10080976>. PubMed PMID.
13. Seni J, Akaro IL, Mkinze B, Kashinje Z, Benard M, Mboowa G, et al. Gastrointestinal tract colonization rate of extended-spectrum beta-lactamase-producing gram-negative Bacteria and Associated factors among Orthopaedic Patients in a Tertiary Hospital in Tanzania: implications for infection Prevention. *Infect Drug Resist*. 2021;14:1733–45. <https://doi.org/10.2147/IDR.S303860>. PubMed PMID: 34007192.
14. Andrews S. FastQC: a quality control tool for high throughput sequence data. Cambridge, United Kingdom: Babraham Bioinformatics, Babraham Institute; 2010.
15. Ewels P, Magnusson M, Lundin S, Käller M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinf (Oxford England)*. 2016;32(19):3047–8. <https://doi.org/10.1093/bioinformatics/btw354>. Epub 2016/06/18.
16. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30(15):2114–20. <https://doi.org/10.1093/bioinformatics/btu170>. Epub 2014/04/04.
17. Seemann T. tseemann/shovill. 2020.
18. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, et al. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS ONE*. 2014;9(11):e112963.
19. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics (Oxford, England)*. 2014;30(14):2068–9. Epub 2014/03/20. <https://doi.org/10.1093/bioinformatics/btu153>. PubMed PMID: 24642063.
20. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence Alignment/Map format and SAMtools. *Bioinf (Oxford England)*. 2009;25(16):2078–9. <https://doi.org/10.1093/bioinformatics/btp352>. *Bioinformatics*.
21. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinf (Oxford England)*. 2009;25(14):1754–60. <https://doi.org/10.1093/bioinformatics/btp324>.
22. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms. SnpEff Fly (Austin). 2012;6(2):80–92. <https://doi.org/10.4161/fly.19695>.
23. Garrison E, Marth G. Haplotype-based variant detection from short-read sequencing. *arXiv:12073907 [q-bio]*. 2012.

Received: 21 April 2023 / Accepted: 4 September 2023

Published online: 07 September 2023

24. Tan A, Abecasis GR, Kang HMJB. Unified Representation of Genetic Variants. 2015;31(13):2202–4.
25. Criscuolo A, Gribaldo S. BMGE (Block Mapping and gathering with Entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. *BMC Evol Biol*. 2010;10(1):210. <https://doi.org/10.1186/1471-2148-10-210>.
26. Ortiz E. vcf2phylop v1. 5: convert a VCF matrix into several matrix formats for phylogenetic analysis; 2018.
27. Katoh K, Standley DMJM. evolution. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. 2013;30(4):772–80.
28. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol*. 2015;32(1):268–74. <https://doi.org/10.1093/molbev/msu300>. Epub 2014/11/06.
29. McArthur AG, Waglechner N, Nizam F, Yan A, Azad MA, Baylay AJ, et al. Compr Antibiotic Resist Database. 2013;57(7):3348–57.
30. Gupta SK, Padmanabhan BR, Diene SM, Lopez-Rojas R, Kempf M, Landraud L et al. ARG-ANNOT, a new bioinformatic tool to discover antibiotic resistance genes in bacterial genomes. 2014;58(1):212–20.
31. Doster E, Lakin SM, Dean CJ, Wolfe C, Young JG, Boucher C et al. MEGARes 2.0: a database for classification of antimicrobial drug, biocide and metal resistance determinants in metagenomic sequence data. *Nucleic Acids Research*. 2019;48(D1):D561–D9. <https://doi.org/10.1093/nar/gkz1010J> Nucleic Acids Research.
32. Chen L, Zheng D, Liu B, Yang J, Jin Q. *Nucleic Acids Res*. 2016;44(D1):D694–7. <https://doi.org/10.1093/nar/gkv1239>. PubMed PMID: 26578559; PubMed Central PMCID: PMC4702877. VFDB 2016: hierarchical and refined dataset for big data analysis—10 years on Epub 2015/11/19.
33. World Health O. Global action plan on antimicrobial resistance. Geneva: World Health Organization; 2015 2015.
34. Najjuka CF, Kateete DP, Kajumbula HM, Joloba ML, Essack SYJBrn. Antimicrobial susceptibility profiles of *Escherichia coli* and *Klebsiella pneumoniae* isolated from outpatients in urban and rural districts of Uganda. 2016;9(1):1–14.
35. Kateraggga JN, Kantume R, Atuhaire C, Lubowa MN, Ndukuu, JGJBp. Toxicology. Phenotypic expression and prevalence of ESBL-producing Enterobacteriaceae in samples collected from patients in various wards of Mulago Hospital. Uganda. 2015;16(1):14.
36. Andrew B, Kagirita A, Bazira JJJlom. Prevalence of extended-spectrum Beta-Lactamases-Producing Microorganisms in patients admitted at KRRH. Southwest Uganda. 2017;2017:3183076.
37. Seni J, Moremi N, Matee M, Van der Meer F, DeVinney R, Mshana S et al. Preliminary insights into the occurrence of similar clones of extended-spectrum beta-lactamase-producing bacteria in humans, animals and the environment in Tanzania: A systematic review and meta-analysis between 2005 and 2016. 2018;65(1):1–10.
38. Sonda T, Kumburu H, van Zwetselaar M, Alifrangis M, Lund O, Kibiki G et al. Meta-analysis of proportion estimates of extended-spectrum-Beta-lactamase-producing Enterobacteriaceae in East Africa hospitals. 2016;5(1):1–9.
39. Jakobsen L, Kuhn KG, Hansen F, Skov RL, Hammerum AM, Littauer PJ et al. Fecal carriage of extended-spectrum and AmpC β -lactamase-producing Enterobacteriaceae in surgical patients before and after antibiotic prophylaxis. 2016;86(3):316–21.
40. Von Wintersdorff CJ, Penders J, Van Niekerk JM, Mills ND, Majumder S, Van Alphen LB, et al. Dissemination of Antimicrobial Resistance in Microbial Ecosystems Through Horizontal gene Transfer. 2016;7:173.
41. Sun D, Jeannot K, Xiao Y, Knapp CW. Horizontal gene transfer mediated bacterial antibiotic resistance. *Front Microbiol*. 2019;10:1933.
42. Al Shaikhli NH, AL OWAIDI A, MARIANA C, MOHSIN S. MOHAMMED HB, GRADISTEANU G. Antibiotic resistance determinants of *Acinetobacter baumannii* strains isolated from nosocomial infections. 2020.
43. Fardini Y, Chettab K, Grépinet O, Rochereau S, Trotereau J, Harvey P, et al. The YfgL lipoprotein is essential for type III secretion system expression and virulence of *Salmonella enterica* Serovar Enteritidis. *Infect Immun*. 2007;75(1):358–70. <https://doi.org/10.1128/iai.00716-06>. Epub 2006/10/25.
44. Seni J, Najjuka CF, Kateete DP, Makobore P, Joloba ML, Kajumbula H et al. Antimicrobial resistance in hospitalized surgical patients: a silently emerging public health concern in Uganda. 2013;6(1):1–7.
45. Seni J, Mapunjo SG, Wittenauer R, Valimba R, Stergachis A, Werth BJ et al. Antimicrobial use across six referral hospitals in Tanzania: a point prevalence survey. 2020;10(12):e042819.
46. Tumbarello M, Spanu T, Di Bidino R, Marchetti M, Ruggeri M, Trecarichi EM, et al. Costs of bloodstream infections caused by *Escherichia coli* and influence of extended-spectrum- β -lactamase production and inadequate initial antibiotic therapy. *Antimicrob Agents Chemother*. 2010;54(10):4085–91.
47. Vihta K-D, Stoesser N, Llewellyn MJ, Quan TP, Davies T, Fawcett NJ, et al. Trends over time in *Escherichia coli* bloodstream infections, urinary tract infections, and antibiotic susceptibilities in Oxfordshire, UK, 1998–2016: a study of electronic health records. *Lancet Infect Dis*. 2018;18(10):1138–49.
48. Mamani R, Flament-Simon SC, García V, Mora A, Alonso MP, López C, et al. Sequence types, Clonotypes, Serotypes, and virotypes of extended-spectrum β -Lactamase-producing *Escherichia coli* Causing Bacteraemia in a Spanish Hospital over a 12-Year period (2000 to 2011). *Front Microbiol*. 2019;10. <https://doi.org/10.3389/fmicb.2019.01530>.
49. Hojabri Z, Mirmohammadvand M, Kamali F, Ghassemi K, Taghavipour S, Pajand O. Molecular epidemiology of *Escherichia coli* sequence type 131 and its H30/H30-Rx subclones recovered from extra-intestinal infections: first report of OXA-48 producing ST131 clone from Iran. *Eur J Clin Microbiol Infect Dis*. 2017;36(10):1859–66. <https://doi.org/10.1007/s10096-017-3021-9>.
50. Liu L, Feng Y, Zhang X, McNally A, Zong Z. New variant of mcr-3 in an extensively drug-resistant *Escherichia coli* clinical isolate carrying mcr-1 and bla NDM-5. *Antimicrob Agents Chemother*. 2017;61(12):e01757–17.
51. Algammal A, Hetta HF, Mabrok M, Behzadi P, Editorial. Emerging multidrug-resistant bacterial pathogens superbugs: a rising public health threat. *Front Microbiol*. 2023;14:1135614. <https://doi.org/10.3389/fmicb.2023.1135614>. Epub 2023/02/24.
52. Burmeister AR. Horizontal Gene Transfer. *Evol Med Public Health*. 2015;2015(1):193–4. <https://doi.org/10.1093/emph/eov018>. PubMed PMID: 26224621.
53. Emamaliipour M, Seidi K, Zununi Vahed S, Jahanban-Esfahlan A, Jaymand M, Majdi H, et al. Horizontal Gene Transfer: from Evolutionary Flexibility to Disease Progression. 2020;8(229). <https://doi.org/10.3389/fcell.2020.00229>.
54. Sevillya G, Adato O, Snir S. Detecting horizontal gene transfer: a probabilistic approach. *BMC Genomics*. 2020;21(1):106. <https://doi.org/10.1186/s12864-019-6395-5>.
55. Woods LC, Gorrell RJ, Taylor F, Connallon T, Kwok T, McDonald MJ. Horizontal gene transfer potentiates adaptation by reducing selective constraints on the spread of genetic variation. *PNAS*. 2020;117(43):26868. <https://doi.org/10.1073/pnas.2005331117>.
56. Boto L. Horizontal gene transfer in evolution: facts and challenges. *Proceedings of the Royal Society B: Biological Sciences*. 2010;277(1683):819–27. <https://doi.org/10.1098/rspb.2009.1679>.
57. Ragupathi NKD, Bakthavatchalam YD, Mathur P, Pragasam AK, Walia K, Ohri VC, et al. Plasmid profiles among some ESKAPE pathogens in a tertiary care centre in south India. *Indian J Med Res*. 2019;149(2):222–31. https://doi.org/10.4103/ijmr.IJMR_2098_17. Epub 2019/06/21.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.