


RESEARCH ARTICLE OPEN ACCESS

Exploring the Genetic Diversity, Virulence and Antimicrobial Resistance of Diarrhoeagenic *Escherichia coli* From Southern Africa Using Whole-Genome Data

Josphat Gichure¹  | Tina Hald² | Elna Buys¹

¹Department of Consumer and Food Sciences, University of Pretoria, Pretoria, South Africa | ²National Food Institute, Technical University of Denmark, Lyngby, Denmark

Correspondence: Elna Buys (elna.buys@up.ac.za)

Received: 14 February 2025 | **Revised:** 6 June 2025 | **Accepted:** 24 July 2025

Funding: The authors acknowledge FOCAL (Foodborne Disease Epidemiology, Surveillance, and Control in African LMIC) Project, a multi-partner, multi-study research grant co-funded by the Bill & Melinda Gates Foundation and the Foreign, Commonwealth & Development Office (FCDO) of the United Kingdom Government [Grant Agreement Investment ID OPP1195617].

Keywords: antibiotic resistance | diarrhoeagenic *Escherichia coli* | Mozambique | virulence | whole-genome sequencing

ABSTRACT

Introduction: Previous studies, including our research, provide critical insights on the contamination of food, water and environment in the Southern African Development Community (SADC) with diarrhoeagenic *Escherichia coli* (DEC). This study used whole-genome sequencing to investigate the genetic diversity, virulence-associated factors and antimicrobial resistance (AMR) patterns of DEC isolated from children under 5 years old and food sources in Maputo and compared these findings with publicly available DEC genome assemblies from the Southern Africa region.

Methods: Whole-genome sequence data from 11 DEC isolates from food, children under 5 and water sources in Maputo, Mozambique, were analysed alongside 125 publicly available DEC genomic assemblies from the SADC region. The latter were retrieved from the Enterobase database (<http://enterobase.warwick.ac.uk>) and included isolates previously collected from food, animals and environmental sources. Genomic analyses were performed using the online pipelines provided by the Centre for Genomic Epidemiology (CGE), Denmark. Unsupervised hierarchical clustering was applied to visualize patterns in genetic diversity, AMR, virulence-associated genes and plasmid content using the R software.

Results: Clustering based on single nucleotide polymorphism (SNP) and core genome multilocus sequence typing (cgMLST) alleles revealed associations based on geographic locations, sample niche, pathovar and O:H antigen, pointing to evolutionary relatedness between the clades with principal coordinate analysis uncovering this accounted for 27.55% of the genetic diversity. Virulence-associated genes encoding for attaching and effacing (*eae*) (63.97%), heat-labile toxin (LT) (25.00%) and Shiga toxin 1 (Stx1) (15.44%) were most abundant, with an inverse association between genes encoding for the presence of LT and *eae*. Resistance to folate pathway antagonists (sulfamethoxazole—55.9%), β -lactamases (amoxicillin, ampicillin and piperacillin—all 54.4%) and aminoglycoside (streptomycin—55.1%) was most abundant.

Conclusions: The study revealed region-specific lineages, evidence of horizontal gene transfer and the clustering patterns suggest both localized and cross-border transmission. The study provides insightful evidence on DEC transmission patterns associated with antimicrobial and disinfectant resistance and associated virulence factors.

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2025 The Author(s). *Public Health Challenges* published by John Wiley & Sons Ltd.

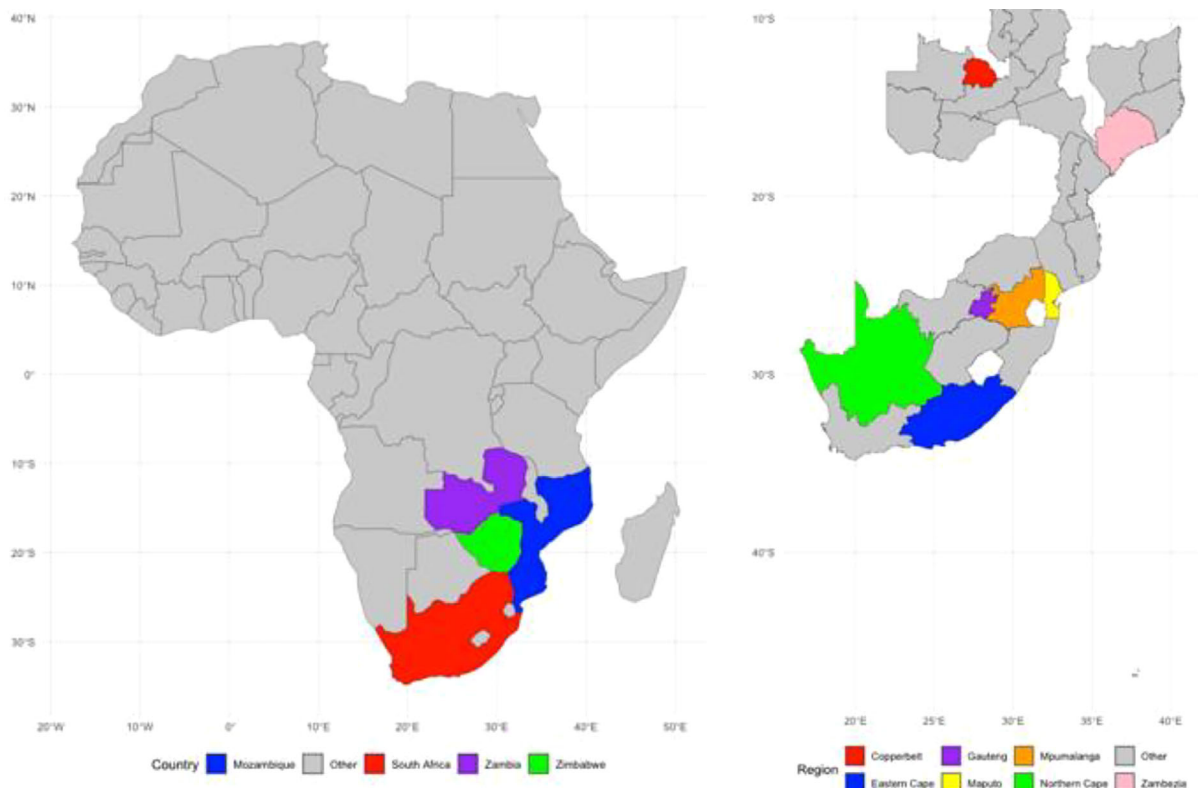


FIGURE 1 | Part (A) highlights the map of Africa with countries that had DEC assemblies publicly available in the EnteroBase database. Part (B) highlights the regions within the countries.

1 | Introduction

Diarrhoeagenic *Escherichia coli* (DEC) remains a significant cause of acute gastroenteritis in the Southern African Development Community (SADC) [1, 2]. Previous studies, including our research (the Foodborne Disease Epidemiology, Surveillance and Control in African LMIC-FOCAL project in Mozambique, Tanzania, Ethiopia and Nigeria), have provided critical insights into the contamination of food and water consumed by children, uncovering the dynamics that influence the transmission of DEC [3, 4]. Despite these advances, questions regarding genetic diversity, antimicrobial resistance (AMR) and virulence factors remain unresolved to expound our understanding of how genetic diversity correlates with source attribution and transmission in Mozambique. This calls for using comparative data from SADC to understand local bacterial evolution, resistance mechanisms and virulence factors. Growing concerns about multidrug resistance (MDR) and extended-spectrum β -lactamases (ESBLs) producing DEC pose a public health risk [5, 6]. Limited local genomic data hinder the development of tailored diagnostic tools and practical treatment guidelines in SADC.

Whole-genome sequencing (WGS) has improved surveillance of pathogens in point-source outbreak investigation, tracking the spread of virulence and antimicrobial-associated resistance genes to the food and environment. Surveillance studies often lack predefined labels or groupings, making detection of emerging trends or novel variants difficult. Unsupervised hierarchical clustering and *k*-means of WGS data have uncovered intrinsic genetic patterns, grouping isolates based on genetic similarity

without prior knowledge. This enables the detection of natural subpopulations that may correspond to distinct lineages, outbreaks or resistance patterns [7, 8].

Despite the significant burden of diarrhoeal disease in Southern Africa, there is limited genomic data on DEC circulating in the region, hindering effective public health interventions. This study addressed a critical gap by analysing the genetic diversity, virulence-associated determinants and AMR profiles of DEC isolates using whole-genome sequence data from children under 5 years old who had diarrhoea and food sources in Maputo, Mozambique, and compared these with publicly available DEC genomic assemblies from the SADC region previously collected from food, animals and environment. Such insights are essential for understanding transmission dynamics, identifying locally circulating lineages and detecting emerging resistance threats. This regional context ensures that data from Mozambique—particularly on unique DEC genetic markers and resistance genes found in Maputo—provides a foundational resource to support antimicrobial stewardship, outbreak preparedness and regional collaboration in managing DEC-related public health threats across the SADC region.

2 | Methodology

2.1 | WGS Data Sources

Eleven DEC isolates, human (7) and food (4), obtained from Maputo, Mozambique, were collected and confirmed as described

A_SNP_Tree scale: 0.01

B_cgMLST_Tree scale: 1000

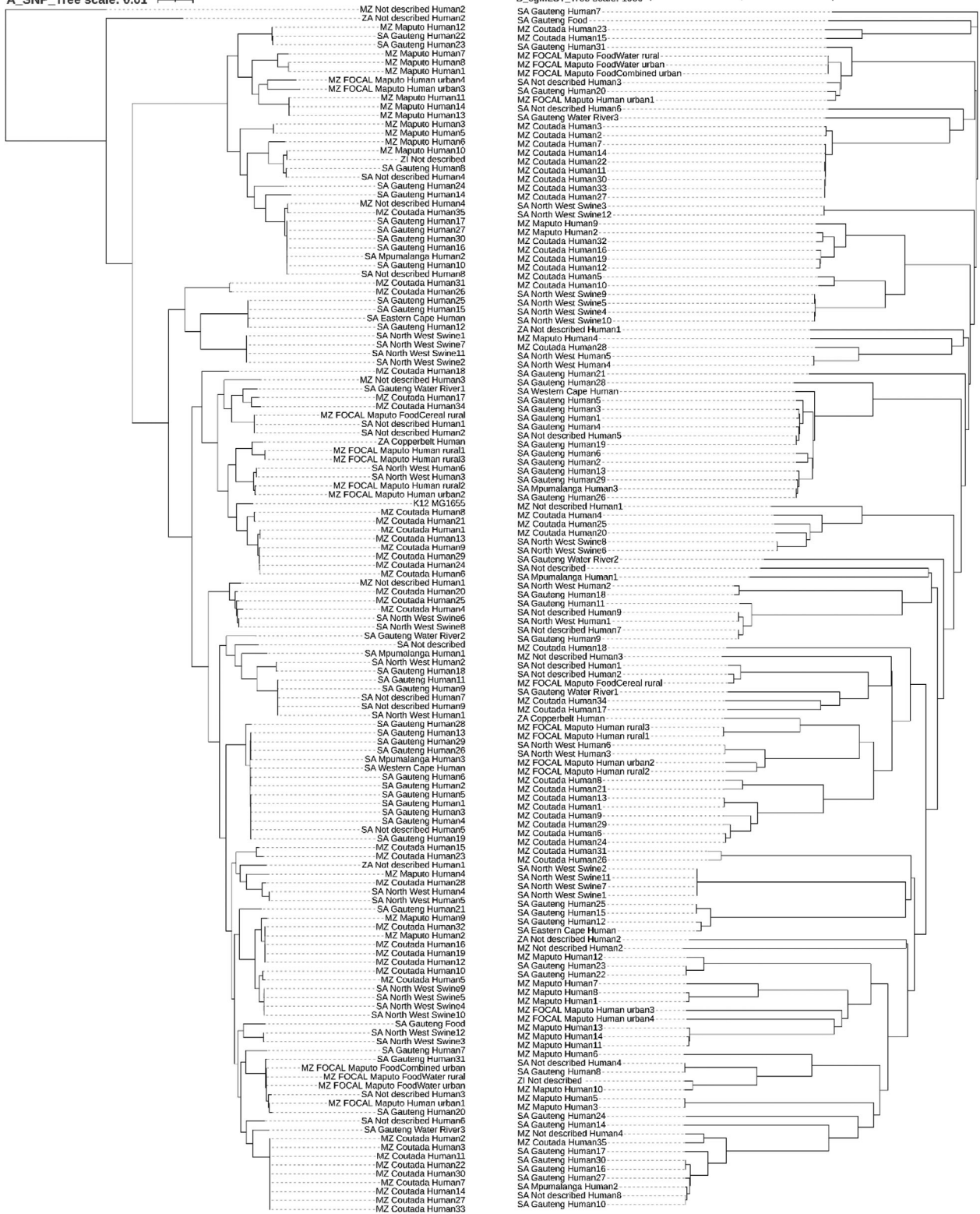


FIGURE 2 | Maximum likelihood phylogenetic trees using SNP (A) and using cgMLST (B) of DEC from Southern Africa. The names have three parts: ZA = Zambia, MZ = Mozambique, SA = South Africa and MZ_FOCAL = the Mozambique FOCAL project. The second part of the names = the provinces (regions) within these countries. The third part = the source of the samples. cgMLST, core genome multilocus sequence typing; DEC, diarrhoeagenic *Escherichia coli*; SNP, single nucleotide polymorphism.

PCoA Plot of cgMLST alleles of E coli from Mozambique and South Africa

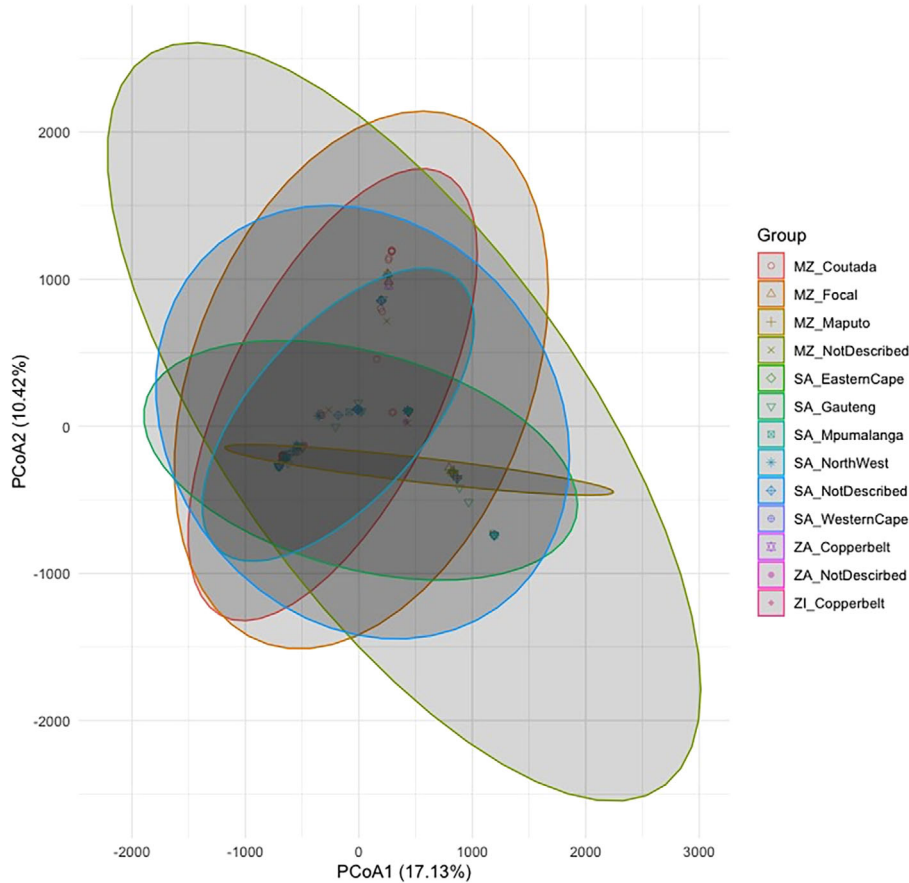


FIGURE 3 | Principal coordinate analysis of cgMLST alleles of 136 DEC strains from the Southern Africa region based on Hamming distance. The names have two parts: ZA = Zambia, MZ = Mozambique, SA = South Africa and MZ_FOCAL = the Mozambique FOCAL project. The second part of the names = the provinces (regions) within these countries. PCoA, principal coordinate analysis; cgMLST, multilocus sequence typing.

in our previous research (Table S1) [3, 4]. WGS was done with the MGI DNBSEQ-G400 sequencing instrument using the MGIEasy Universal DNA Library Prep Kit (https://en.mgi-tech.com/products/reagents_info/id/8) protocol at the Agricultural Research Council (Onderstepoort, Pretoria, South Africa). The non-interleaved paired-end FASTQ sequence files were uploaded to Kbase (<https://www.kbase.us/>), an online web-based platform [9]. FastQC (v0.12.1) was used to assess sequence file quality, whereas trimming and removal of the adapters from the reads were done using Trimmomatic (v0.36). De novo assembly was used to align the short reads into contigs using SPAdes (v3.15.3) [10]. The non-interleaved paired-end FASTQ sequence files were also uploaded to Enterobase v1.2.0 (<http://enterobase.warwick.ac.uk>) database web-based platform [11]. The presence of virulence-associated factors encoding for Shiga toxin 1 (Stx1), Shiga toxin 2 (Stx2), invasion plasmid antigen H (*ipaH*), invasion plasmid (pInv), heat-stable enterotoxin (ST), heat-labile enterotoxin (LT) or intimin (*eae*) was used to confer the pathovar. One hundred and twenty-five DEC assemblies from the Southern Africa region that were publicly available in the Enterobase database were also obtained. The metadata comprising the uber-strain number, assembly barcode, source details, collection year, rST, MLST, core genome multilocus sequence typing (cgMLST),

wgMLST, pathovar, Clermont Typing and O:H antigen of the 136 DEC isolates analysed in this study has been attached in the Supporting Information section. The assemblies, single nucleotide polymorphism (SNP) matrix and alleles obtained from the Enterobase database were downloaded for downstream processing. The map of the Southern Africa region from which the DEC assemblies were obtained is presented in Figure 1A, whereas Figure 1B highlights the regions within these countries. The unknown regions have not been included, whereas Coutada 11 is in Zambezia province in Mozambique.

2.2 | Analysis of Genome

The assemblies of the DEC isolates were analysed using a web-based pipelines at the Centre for Genomic Epidemiology (CGE), DTU, Denmark (<https://www.genomicepidemiology.org/services/>). Core genome alleles were obtained using cgMLSTFinder 1.2 [11, 12]. ResFinder 4.5.0 was used to identify acquired antimicrobial and disinfectant resistance genes and chromosomal point mutations based on the following genes: *gyrA*, *gyrB*, *parC*, *parE*, *pmrA*, *pmrB*, *folP*, 23S, 16S-*rrsB*, 16S-*rrsC*, 16S-*rrsH*, *ampC*-promoter-size-53 bp, *rpoB* [13, 14]. PlasmidFinder 2.1 was

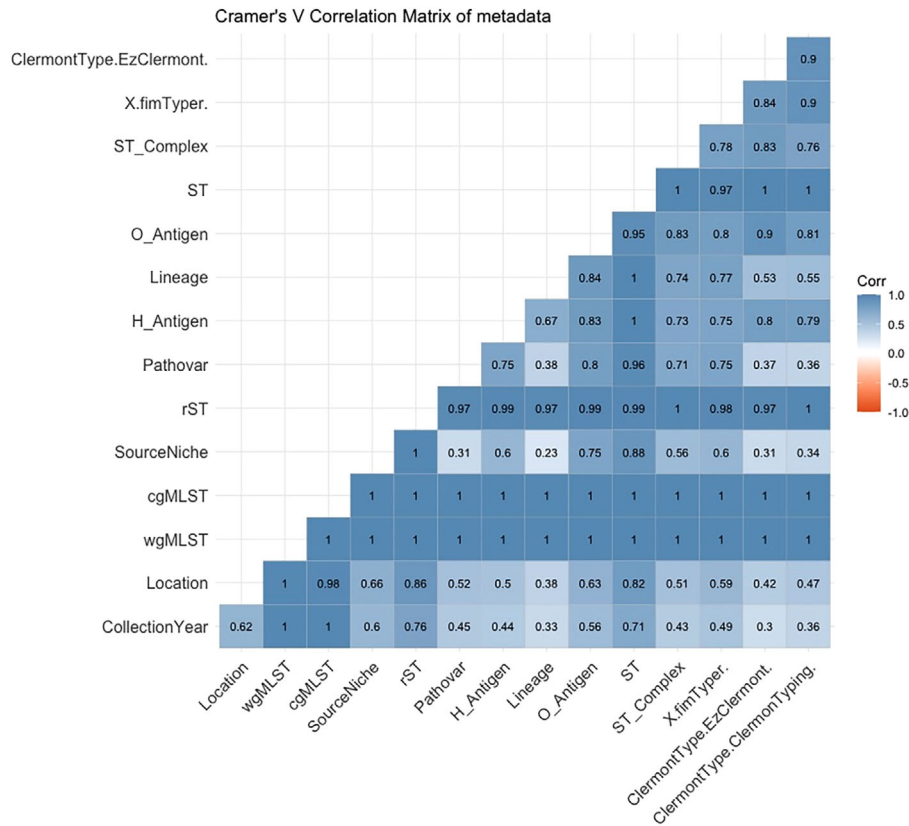


FIGURE 4 | The correlation matrix of metadata of WGS data of DEC isolates alleles from the Southern Africa region. cgMLST, core genome multilocus sequence typing.

used to identify plasmids that confer antibiotic and disinfectant resistance [14, 15].

E. coli str. K-12 substr. MG1655 obtained from the NCBI website was used as the reference genome to confer phylogeny as recommended [16]. SNP analysis and identification of the phylogenies, MLST, cgMLST, rMLST and serotypes were performed using EnteroBase with all the parameters set at default [11]. Hamming distance was used to perform a pairwise distance matrix between the cgMLST alleles with the unweighted pair group method with arithmetic mean (UPGMA) algorithm used to perform hierarchical clustering [17] in R with 100 bootstraps to enhance the reliability of the tree nodes. Clustering the SNP was performed using the Neighbour-Joining algorithm [17]. For MLST, alleles of seven housekeeping genes, *adhA*, *fumC*, *gyrB*, *icd*, *mdh*, *purA* and *recA*, were used. The serotype of the *E. coli* genome was performed using EnteroBase, based on the O-antigen (lipopolysaccharide) and H-antigen (flagellar) genes [18].

The alleles and annotations of the core genomes were analysed using R statistical environment (V 4.2.2), ape (v 5.6-2), phangorn (v 2.11.1), pvclust (v 2.2-0) and compGenomeRData (v 0.1.0), whereas visualization was done using dendextend (v 1.17.1) packages.

3 | Results and Discussion

3.1 | Genetic Diversity Clustering Patterns

3.1.1 | Phylogenetic Analysis on SNP and cgMLST

A side-by-side visual comparison of dendrograms generated using SNP and cgMLST amongst the 136 DEC isolates from the Southern Africa region has been presented in Figure 2A using SNP matrix and Figure 2B using cgMLST. For SNP, 341,044 variant sites and 3068 accessory genomes were identified, with the minimum percentage sites maintained at 95%, whereas 2513 alleles of core genomes were identified for cgMLST. There was evidence of isolates from similar geographical locations clustering revealing localized transmission patterns evolution. The side-by-side comparison of SNP and cgMLST dendrograms clustered the isolates with a fairly similar trend observed.

A complete description and STs of the DEC isolates with the metadata have been presented in the Supporting Information section (Supplementart Data 2, Supplementart Data 1, Figure S1). Sixty-three unique STs were identified from which ST 21 ($n = 13$, 9.56%), ST 443 ($n = 9$, 6.62%), ST 583 ($n = 8$, 5.88%) and ST 40 ($n = 7$, 5.15%) were most abundant (Figure 4). The isolates were further categorized on the basis of the ST complex (Cplx), where 97 were

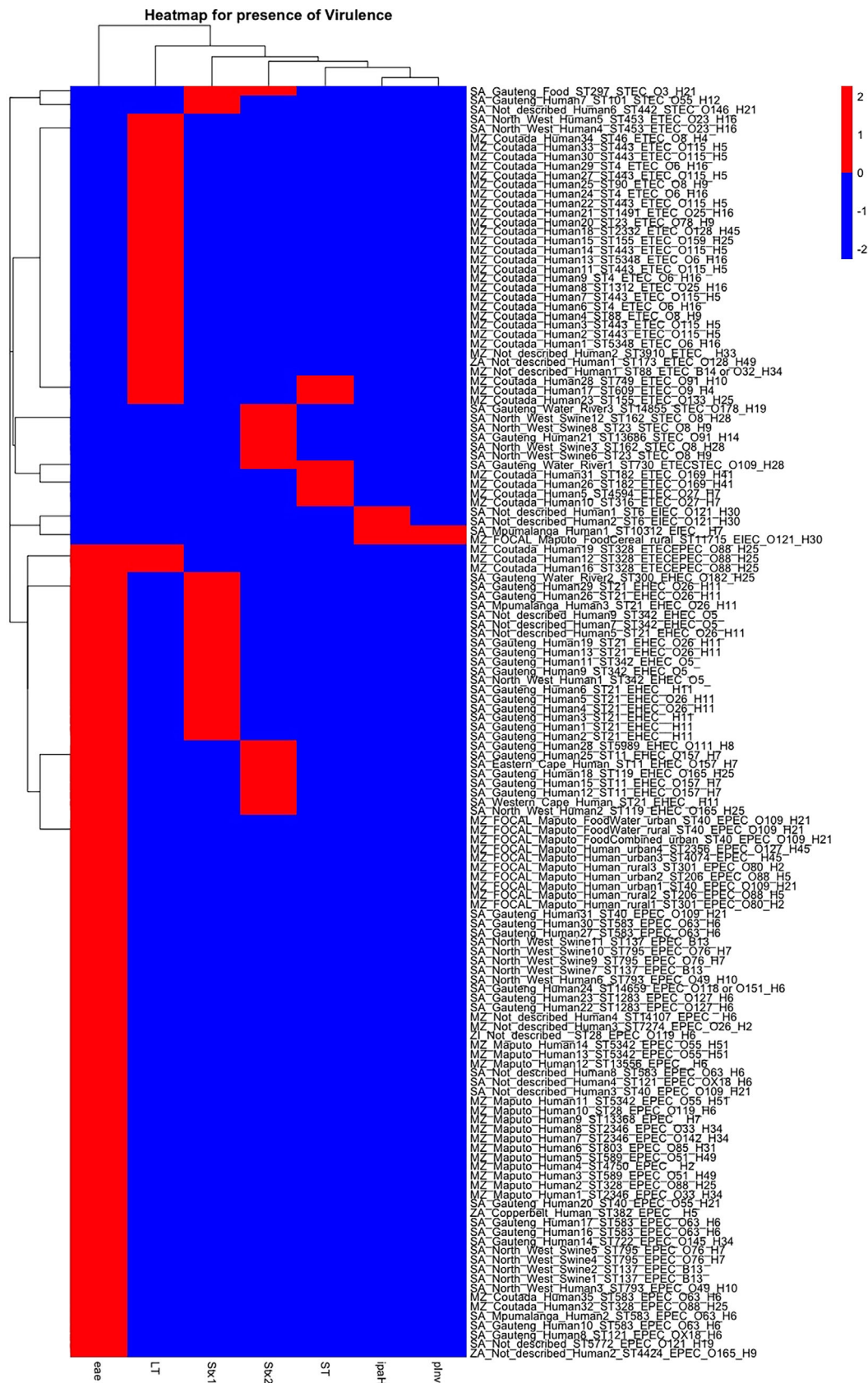


FIGURE 5 | Heatmap on the presence of virulence factors for DEC isolates in Southern Africa. The names have five parts: ZA = Zambia, MZ = Mozambique, SA = South Africa and MZ_FOCAL = the Mozambique FOCAL project. The second part of the names = the provinces (regions) within these countries. The third part = the source of the samples. The fourth = the ST number. The fifth part = the pathotype. The sixth part = the O:H antigen.

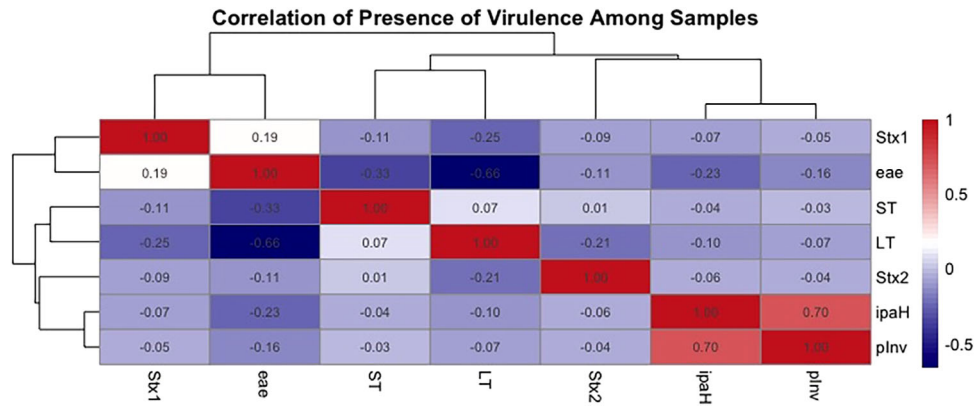


FIGURE 6 | Correlation values on the presence of virulence factors among DEC isolates from the Southern Africa region.

TABLE 1 | Chromosomal point mutation conferring antimicrobial resistance of diarrhoeagenic *Escherichia coli* (DEC) isolates from the Southern Africa region.

	Sample	Gene	Mutation	Resistance	PubMed identifier
1	ZA_Copperbelt_Human	<i>gyrA</i>	<i>gyrA</i> p.S83L	Nalidixic acid, Ciprofloxacin	8891148
2	MZ_Maputo_Human1	<i>gyrA</i>	<i>gyrA</i> p.S83A	Nalidixic acid, Ciprofloxacin	12654733
3	ZA_Not_described_Human2	<i>parE</i>	<i>parE</i> p.I355T	Nalidixic acid, Ciprofloxacin	28598203
4	SA_Gauteng_Human22	<i>gyrA</i>	<i>gyrA</i> p.S83L	Nalidixic acid, Ciprofloxacin	8891148
5	SA_Gauteng_Human23	<i>gyrA</i>	<i>gyrA</i> p.S83L	Nalidixic acid, Ciprofloxacin	8891148
6	SA_North_West_Human6	<i>parC</i>	<i>parC</i> p.A56T	Nalidixic acid, Ciprofloxacin	12654733
7	MZ_FOCAL_Maputo_Human_rural2	<i>parC</i>	<i>parC</i> p.A56T	Nalidixic acid, Ciprofloxacin	12654733
8	MZ_FOCAL_Maputo_Human_urban2	<i>parC</i>	<i>parC</i> p.A56T	Nalidixic acid, Ciprofloxacin	12654733
9	MZ_FOCAL_Maputo_Human_rural1	<i>gyrA</i>	<i>gyrA</i> p.S83L	Nalidixic acid, Ciprofloxacin	8891148
		<i>parC</i>	<i>parC</i> p.S80I	Nalidixic acid, Ciprofloxacin	8851598
10	MZ_FOCAL_Maputo_Human_rural3	<i>gyrA</i>	<i>gyrA</i> p.S83L	Nalidixic acid, Ciprofloxacin	8891148
		<i>parC</i>	<i>parC</i> p.S80I	Nalidixic acid, Ciprofloxacin	8851598
11	MZ_Not_described_Human2	<i>parE</i>	<i>parE</i> p.I355T	Nalidixic acid, Ciprofloxacin	28598203
12	SA_North_West_Human3	<i>parC</i>	<i>parC</i> p.A56T	Nalidixic acid, Ciprofloxacin	12654733

Note: The names have three parts: ZA = Zambia, MZ = Mozambique, SA = South Africa and MZ_FOCAL = our current FOCAL project. The second part = the provinces (regions) within these countries. The third part = the sources.

Heatmap of Resistance to AMR and Disinfectant Abundance

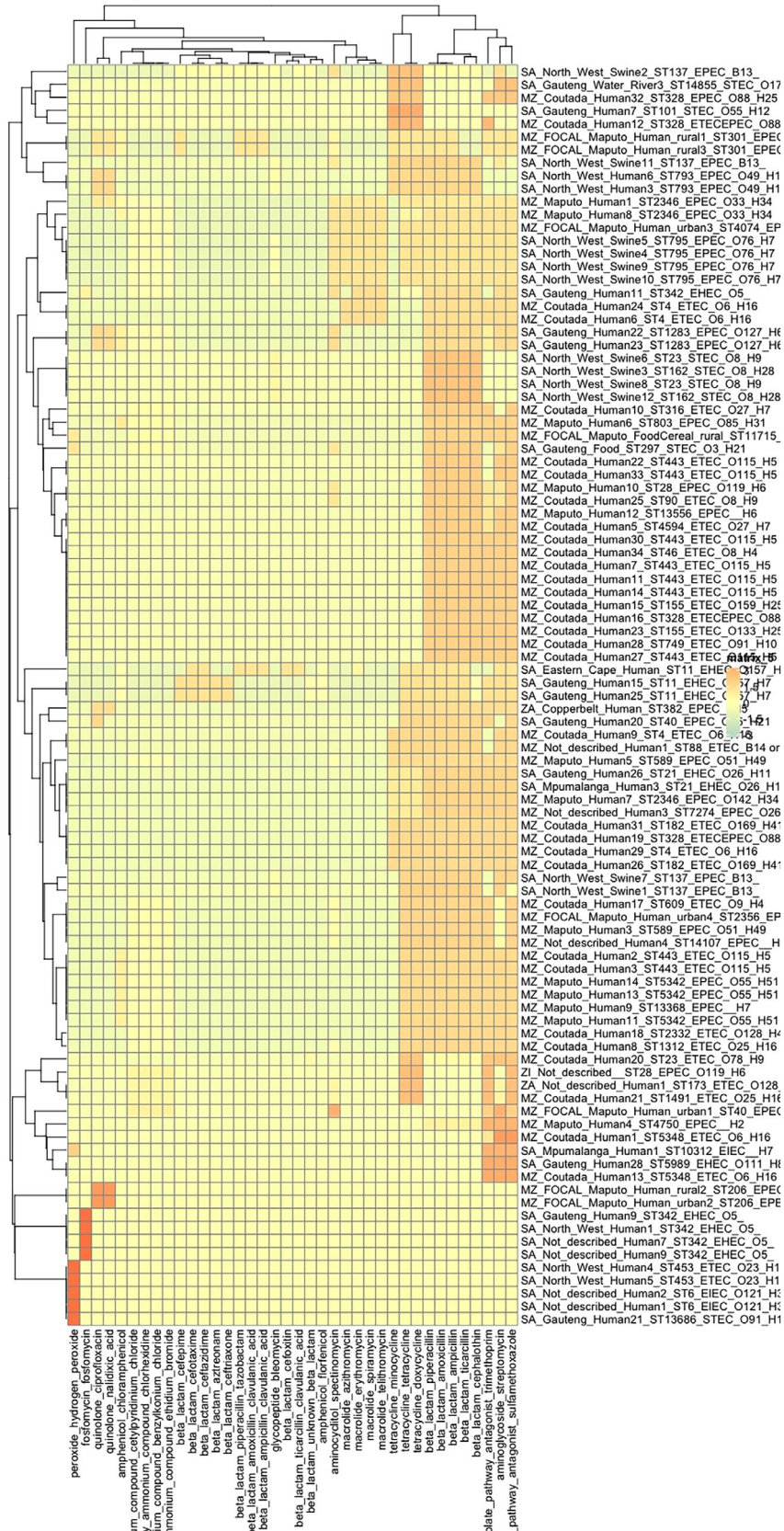


FIGURE 7 | Heatmap of resistance to antimicrobial and disinfectants among DEC isolates from Southern Africa region. The names have five parts: ZA = Zambia, MZ = Mozambique, SA = South Africa and MZ_FOCAL = the Mozambique FOCAL project. The second part of the names = the provinces (regions) within these countries. The third part = the source of the samples. The fourth = the ST number. The fifth part = the pathotype. The sixth part = the O:H antigen.

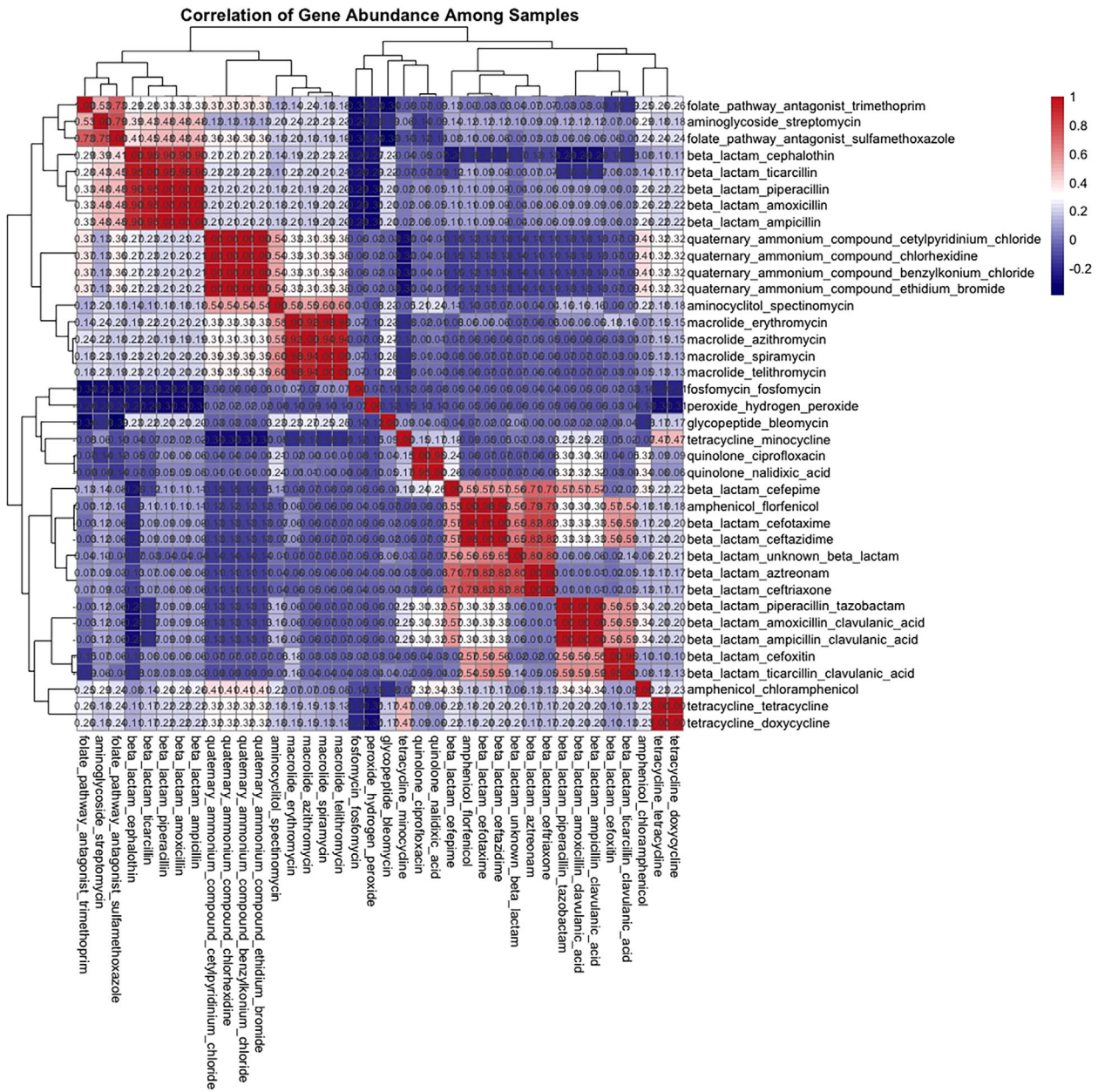


FIGURE 8 | Correlation matrix of the abundance patterns of resistance to various antimicrobials and disinfectants among the DEC isolates.

identified, whereas 39 were unknown. ST 29 Cplx ($n = 14$, 14.43%), ST 278 Cplx ($n = 12$, 12.37%), ST 205 Cplx ($n = 10$, 10.31%), ST 122 Cplx ($n = 9$, 9.28%) and ST 10 Cplx ($n = 8$, 8.25%) are the most abundant complexes. On the basis of rMLST, 77 unique profiles with rST 2264 ($n = 11$), rST 2199 ($n = 9$), rST 1523 ($n = 8$) and rST 14908 ($n = 4$) are the most abundant. On the basis of cgMLST, 121 unique cgSTs were reported, with 15 cgST duplicates and none appearing more than twice. All the 136 wgMLST profiles were unique.

The phylogenetic relationships based on SNPs amongst the 136 DEC isolates based on hierarchical clustering using the UPGMA algorithm, where each point represents an isolate and colours indicate cluster membership, have been presented in Figure S2. The optimal number of k -means nine clusters determined using the Elbow method has been presented in Figure S3. The

number of DEC isolates in each cluster has been presented in Figure S2. The hierarchical relationships of 136 DEC isolates from the Southern Africa region based on their cgMLST alleles profiles have been presented in Figure S4. The p values of each branch within the clusters have been presented in Figure S5. The bootstrap results on the consistency of cgMLST cluster assignment have been presented in Figure S6.

From both algorithms, the clusters were predominantly associated with geographic region, MLST, pathovar and O–H antigen, pointing to potential epidemiological linkage. In addition, genetic dissimilarities were observed among samples from different source niches within similar geographical locations. To validate the robustness of k -means clustering, the study performed hierarchical clustering and observed consistent subgroup patterns, reinforcing the reliability of the k -means clusters. K -means

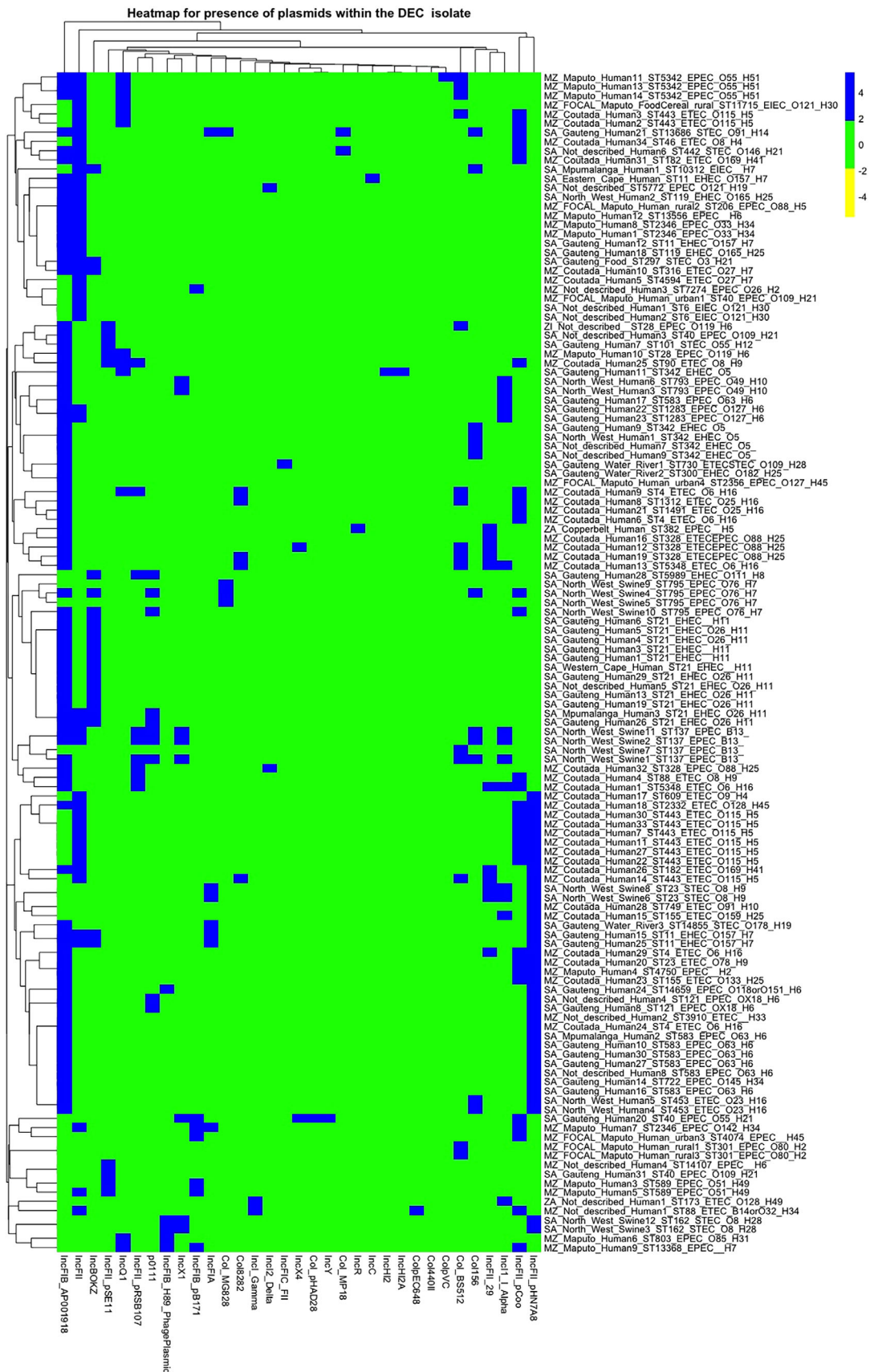


FIGURE 9 | Hierarchical clustering of DEC isolates from the Southern Africa region based on the presence of plasmids. The names have five parts: ZA = Zambia, MZ = Mozambique, SA = South Africa and MZ_FOCAL = Mozambique FOCAL project. The second part of the names = the provinces (regions) within these countries. The third part = the source of the samples. The fourth = the ST number. The fifth part = the pathotype. The sixth part = the O:H antigen. DEC, diarrhoeagenic *Escherichia coli*.

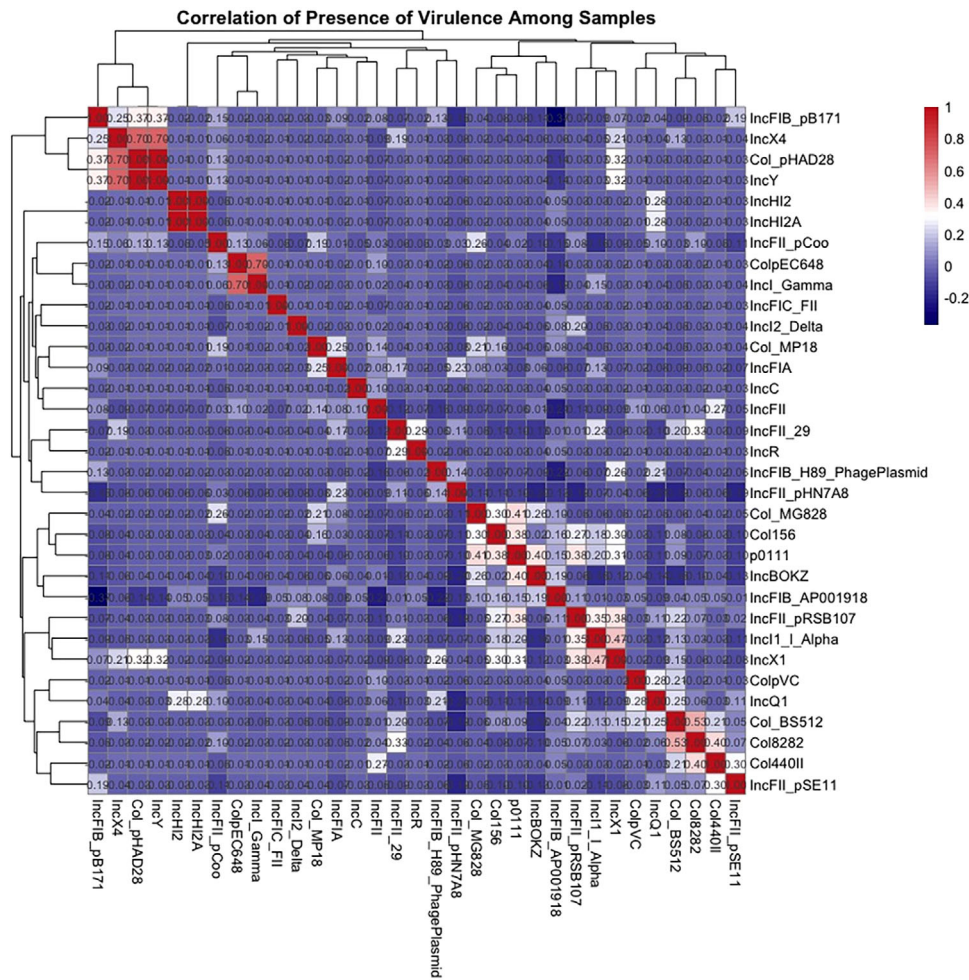


FIGURE 10 | Correlation matrix showing the presence of plasmids among 136 DEC isolates from the Southern Africa region.

clustering assumes spherical clusters; therefore, there is a need to capture the underlying genetic relationships to generalize groupings fully [19]. These findings highlight the need to integrate genetic diversity among DEC isolates with epidemiological surveillance in DEC outbreak investigations.

The principal coordinate analysis (PCoA) of the DEC isolates has been presented in Figure 3. The cgMLST analysis shows that 17.13% and 10.42% of the total genetic variability could be attributed to PCoA1 and PCoA2, respectively. The ellipse further revealed that the genetic diversity of isolates from similar geographical locations tends to cluster together.

From these clustering algorithms, geographical differentiation played a primary role in the clustering of local clones. As reported in other studies, phylogenetic clustering was not affected by epidemiological factors within the region [19].

3.1.2 | Association Between Diversity and Metadata

The strength and significance of associations of variables within the metadata have been presented in Figure 4. The moderate association between the pathovar and the collection year, geographical location and source niche indicates no direct linkage

with pathovars. The presence of local clades was observed as the location had an increasing association as the number of alleles increased, with wgMLST having a solid association, whereas MLST had the least, which indicates that genetic diversity was strongly associated with geographic location.

3.2 | Virulence and Resistance Associated Genes Clustering Patterns of the DEC Isolates

3.2.1 | Virulence Factors Profiling

A heatmap on the presence of virulence-associated genes for DEC isolates from the Southern Africa region is presented in Figure 5. Of the 136 DEC isolates, 36 had two virulence-associated genes, whereas the rest had just one. The most abundant virulence-associated genes encode for *eae* (63.97%), *LT* (25.00%), *Stx1* (15.44%), *Stx2* (11.76%), *ST* (5.88%), *ipaH* (2.94%) and *pInv* (1.47%). Clustering revealed that geographical location was associated with virulence. Fifty-eight of the DEC isolates were EPEC, 35 were ETEC, 26 were EHEC, 9 were STEC, 4 were EIEC, and 3 were ETEC/EPEC, whereas 1 was ETEC/STEC.

The correlation matrix showing the presence of virulence-associated genes among the DEC isolates has been presented

in Figure 6. A negative correlation was observed between the presence of genes encoding for *eae* and *Stx1* and other virulence-associated determinants, whereas presence of genes encoding *ipaH* and *pInv* had a positive relationship. These patterns suggest potential horizontal transfer of virulence genes within DEC clades from similar geographic regions, possibly driven by local selective pressures. Although gene co-occurrence has been documented in AMR studies, comparable data for DEC remain limited. These findings may support the hypothesis of co-virulence, as previously proposed, where the combined presence of multiple virulence factors can enhance the pathogenic potential [20].

Mobile genetic elements such as plasmids, transposons, phages and pathogenicity islands play a key role in virulence [21]. Unlike studies from Europe and North America, there were only 4/136 O157:H7 serotypes, with O26:H11 (9/136), O115:H5 (9/136) and O63:H6 (8/136) being the predominant ones which rhyme existing epidemiological data [22, 23].

3.2.2 | AMR Profiles

A heatmap showing the likelihood and abundance of AMR and disinfectant-associated genes among the 136 DEC isolates has been presented in Figure 7. Of these, 97 carried genes conferring resistance to at least one antimicrobial agent or disinfectant. The clustering revealed that resistance profiles were associated with both the type and abundance of the resistance genes. The proportion of DEC isolates resistance to specific antimicrobials or disinfectants is detailed in Figure S7. Notably, clustering appeared to be driven primarily by the geographical location, whereas the ST number, pathovar and O:H antigen played a lesser role.

The most frequently identified resistance genes were those associated with folate pathway antagonists (sulfamethoxazole—55.9% and trimethoprim (49.3), as well as β -lactams (amoxicillin—54.4%, ampicillin—54.4%, piperacillin—54.4%, ticarcillin—52.9 and cephalothin—52.2%). Additionally, resistance to aminoglycosides (streptomycin—55.1%) and tetracycline (doxycycline and tetracycline, both 38.2%) was prevalent, suggesting the presence of MDR.

Identifying clusters with distinct virulence and resistance profiles highlights the genetic diversity among DEC isolates and underscores the need for targeted surveillance and intervention strategies. The study found resistance to some antibiotics in the reserved category such as aztreonam, ceftazidime and fosfomicin, and some in the watch category such as azithromycin, cefotaxime and piperacillin + tazobactam according to the WHO AWaRe (Access, Watch, Reserve) guidelines (<https://list.essentialmeds.org/antibiotics/watch>).

The presence of MDR and ESBL-producing DEC isolates in the Southern African region poses a public health concern as previously reported [24]. Similar to other studies, AMR to fourth-generation cefepime and third-generation ceftazidime, cefotaxime and ceftriaxome has been previously reported in the Southern Africa region [25–27].

Figure 8 presents the correlation matrix showing the abundance patterns of resistance to various antimicrobials and disinfectants among the DEC isolates. The correlation values ranged from -0.4 to 1 , indicating relationships from weak negative to perfect positive correlations. Notably, resistance to disinfectants such as quaternary ammonium compounds and peroxides had a weak negative correlation to mainly β -lactams antimicrobials.

Only 12 of the 136 DEC isolates had chromosomal point mutation, as presented in Table 1. The identified mutations in the genes *gyrA*, *parC* and *parE* are well-known markers for quinolone resistance (nalidixic acid and ciprofloxacin) which is consistent with the observed resistance profiles. In addition, all the isolates that conferred *gyrA*, *parC* and *parE* mutations were human isolates. Transferred genes are primarily concentrated in about 1% of chromosomal regions, known as hot spots, and their presence highlights extrachromosomal elements in integrating adaptive resistance genes, as suggested [28].

3.2.2.1 Plasmids. A heatmap of plasmid harbouring AMR present in the DEC isolates from the Southern Africa region has been presented in Figure 9. The top 10 most abundant plasmid genes within the DEC isolates have been presented in Figure S8. Eight of the 136 DEC isolates had no plasmids, and of the eight, four were from the current research. IncFIB(AP001918) (23.06%), IncFII (12.78%), IncFII(pHN7A8) (9.27%), IncFII(pCoo) (8.77%) and IncB/O/K/Z (6.02%) were the most abundant plasmid. IncFIA, IncFIB, IncFIC and IncB/O have all been reported to be capable of carrying transfer, MDR and virulence-associated genes and were all observed in this project [29]. The IncFII plasmids have been reported to be abundant among the resistance plasmids in Enterobacteriaceae from different sources, with their occurrence associated with AMR [30]. IncF plasmids have been reported to mobilize high-pathogenicity islands [21]. IncFIB plasmids contribute to resistance to antibiotics such as β -lactams, quinolones, aminoglycosides, tetracyclines and phenicols [31]. The distribution of plasmids among the DEC isolates was influenced by geographical location, suggesting horizontal plasmid transfer within regions.

A correlation matrix of DEC isolates from the Southern Africa region based on the presence of plasmids has been presented in Figure 10. It was observed that Inc-type plasmids generally exhibited a weak negative correlation with other Inc plasmids except for IncH12 and IncH12A. In addition, Inc-type showed weak correlations with Col-type and other non-Inc plasmids. However, notably strong positive correlations were observed among IncX4, Col_pHAD28 and IncY plasmids, as well as ColpEC648 and IncI_Gamma plasmids, indicating they frequently co-occur within the same isolates.

4 | Conclusion

This study explored the genetic diversity, virulence-associated determinants and AMR patterns of DEC isolates from Southern Africa using WGS data. By analysing isolates from foods consumed by children under five and faecal samples from those children in Maputo, alongside publicly available genomic data from the SADC region, this study reveals region-specific genetic patterns, shared lineages and potential transmission pathways.

The observed diversity in virulence-associated factors and antimicrobial and disinfectant resistance highlights the capacity of DEC to adapt and persist across various ecological niches.

By applying unsupervised learning approaches, we identified epidemiologically relevant clusters based on geography, resistance patterns and transmission pathways. Clustering based on rST, MLST, cgMLST, wgMLST and O:H serotypes revealed the presence of closely related clades across different geographic locations, sources and collection periods, suggesting both localized contamination and broader regional dissemination. Evidence of horizontal gene transfer through shared virulence and resistance-associated genes, chromosomal mutations and plasmids further underscores the dynamic evolution of DEC in response to environmental and anthropogenic pressures.

These findings provide critical genomic insights that can inform public health surveillance, outbreak detection and antimicrobial stewardship across Mozambique and the broader Southern African region. Integrating local and regional genomic data is essential to understanding DEC transmission dynamics and guiding effective interventions to reduce its impact on vulnerable populations, particularly young children.

Author Contributions

Josphat Gichure: Conceptualization (equal); Formal analysis (equal); Methodology (equal); Writing original draft (equal); Writing review and editing (equal). **Tine Hald:** Conceptualization (equal); Writing review and editing (supporting); Funding acquisition. **Elna Buys:** Conceptualization (equal); Methodology (supporting); Supervision (lead); Writing review and editing (supporting).

Acknowledgements

The authors acknowledge FOCAL (Foodborne Disease Epidemiology, Surveillance, and Control in African LMIC) Project, a multi-partner, multi-study research grant co-funded by the Bill & Melinda Gates Foundation and the Foreign, Commonwealth & Development Office (FCDO) of the United Kingdom Government [Grant Agreement Investment ID OPP1195617].

Ethics Statement

The DEC isolates collected from Mozambique are from the Foodborne Disease Epidemiology, Surveillance, and Control in African LMIC-FOCAL project. Ethical clearance (registration number: CIBS FM & amp; HCM/092/2019) was obtained from the institutional health bioethics committee at the Faculty of Medicine/Maputo Central Hospital.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The whole-genome sequence (WGS) data generated in this study will be available in the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA). The accession number will be shared, and data will be publicly accessible upon publication. Prior to publication, data can be made available upon reasonable request.

Permission to Reproduce Material From Other Sources

We shall do a written request to Enterobase. We acknowledge that these data were reproduced as per Enterobase's data usage policies.

References

1. M. Amukoshi, I. Maposa, R. S. Moyo, and M. Mukesi, "Etiological Agents Isolated From Stool Samples of Children Under the Age of Five Years in Windhoek, Namibia," *Edorium Journal of Microbiology* 3 (2017): 1–9, <https://doi.org/10.5348/M08-2017-6-OA-1>.
2. T. B. Oppong, H. Yang, C. Amponsem-Boateng, et al., "Enteric Pathogens Associated With Gastroenteritis Among Children Under 5 Years in Sub-Saharan Africa: A Systematic Review and Meta-Analysis," *Epidemiology and Infection* 148, no. e64 (2020): 1–9, <https://doi.org/10.1017/S0950268820000618>.
3. B. N. Desta, T. Gobena, C. Macuamule, et al., "Practicalities of Implementing Burden of Disease Research in Africa: Lessons From a Population Survey Component of Our Multi-Partner FOCAL Research Project," *Emerging Themes in Epidemiology* 19, no. 1 (2022): 1–12, <https://doi.org/10.1186/s12982-022-00113-y>.
4. S. Faife, C. Macuamule, J. Gichure, T. Hald, and E. Buys, "Diar-rhoeagenic *Escherichia coli* and *Salmonella* spp. Contamination of Food and Water Consumed by Children With Diarrhoea in Maputo, Mozambique," *International Journal of Environmental Research and Public Health* 21, no. 9 (2024): 1122, <https://doi.org/10.3390/ijerph21091122>.
5. J. Mbangi, D. G. Amoako, A. L. K. Abia, M. Allam, A. Ismail, and S. Y. Essack, "Genomic Insights of Multidrug-Resistant *Escherichia coli* From Wastewater Sources and Their Association With Clinical Pathogens in South Africa," *Frontiers in Veterinary Science* 8 (2021): 1–13, <https://doi.org/10.3389/fvets.2021.636715>.
6. T. Ramatla, M. Tawana, K. E. Lekota, and O. Thekisoe, "Antimicrobial Resistance Genes of *Escherichia coli*, a Bacterium of "One Health" Importance in South Africa: Systematic Review and Meta-Analysis," *AIMS Microbiology* 9, no. 1 (2023): 75–89, <https://doi.org/10.3934/microbiol.2023005>.
7. J. Herrero, A. Valencia, and J. Dopazo, "A Hierarchical Unsupervised Growing Neural Network for Clustering Gene Expression Patterns," *Bioinformatics* 17, no. 2 (2001): 126–136, <https://doi.org/10.1093/bioinformatics/17.2.126>.
8. R. Mukherjee, J. Vidic, S. Auger, H. C. Wen, R. P. Pandey, and C. M. Chang, "Exploring Disease Management and Control Through Pathogen Diagnostics and One Health Initiative: A Concise Review," *Antibiotics* 13, no. 1 (2024): 1–16, <https://doi.org/10.3390/antibiotics13010017>.
9. B. Allen, M. Drake, N. Harris, and T. Sullivan, "Using KBase to Assemble and Annotate Prokaryotic Genomes," *Current Protocols in Microbiology* 46 (2017): 1E.13.1–1E.13.18, <https://doi.org/10.1002/cpmc.37>.
10. A. Bankevich, S. Nurk, D. Antipov, et al., "SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing," *Journal of Computational Biology* 19, no. 5 (2012): 455–477, <https://doi.org/10.1089/cmb.2012.0021>.
11. Z. Zhou, N. F. Alikhan, K. Mohamed, Y. Fan, and M. Achtman, "The Enterobase User's Guide, With Case Studies on *Salmonella* Transmissions, *Yersinia pestis* Phylogeny, and *Escherichia* Core Genomic Diversity," *Genome Research* 30, no. 1 (2020): 138–152, <https://doi.org/10.1101/gr.251678.119>.
12. P. T. L. C. Clausen, F. M. Aarestrup, and O. Lund, "Rapid and Precise Alignment of Raw Reads Against Redundant Databases With KMA," *BMC Bioinformatics [Electronic Resource]* 19, no. 1 (2018): 1–8, <https://doi.org/10.1186/s12859-018-2336-6>.
13. V. Bortolaia, R. S. Kaas, E. Ruppe, et al., "ResFinder 4.0 for Predictions of Phenotypes From Genotypes," *Journal of Antimicrobial Chemotherapy* 75, no. 12 (2020): 3491–3500, <https://doi.org/10.1093/jac/dkaa345>.

14. C. Camacho, G. Coulouris, V. Avagyan, et al., “BLAST+: Architecture and Applications,” *BMC Bioinformatics [Electronic Resource]* 10 (2009): 1–9, <https://doi.org/10.1186/1471-2105-10-421>.
15. A. Carattoli, E. Zankari, A. García-Fernández, et al., “In Silico Detection and Typing of Plasmids Using Plasmidfinder and Plasmid Multilocus Sequence Typing,” *Antimicrobial Agents and Chemotherapy* 58, no. 7 (2014): 3895–3903, <https://doi.org/10.1128/AAC.02412-14>.
16. N. A. O’Leary, M. W. Wright, J. R. Brister, et al., “Reference Sequence (RefSeq) Database at NCBI: Current Status, Taxonomic Expansion, and Functional Annotation,” *Nucleic Acids Research* 44, no. D1 (2016): D733–D745, <https://doi.org/10.1093/nar/gkv1189>.
17. C. Vaz, M. Nascimento, J. A. Carriço, T. Rocher, and A. P. Francisco, “Distance-Based Phylogenetic Inference From Typing Data: A Unifying View,” *Briefings in Bioinformatics* 22, no. 3 (2021): 1–16, <https://doi.org/10.1093/bib/bbaa147>.
18. K. G. Joensen, A. M. M. Tetzschner, A. Iguchi, F. M. Aarestrup, and F. Scheutz, “Rapid and Easy In Silico Serotyping of *Escherichia coli* Isolates by Use of Whole-Genome Sequencing Data,” *Journal of Clinical Microbiology* 53, no. 8 (2015): 2410–2426, <https://doi.org/10.1128/JCM.00008-15>.
19. A. M. Ikotun, A. E. Ezugwu, L. Abualigah, B. Abuhaija, and J. Heming, “K-means Clustering Algorithms: A Comprehensive Review, Variants Analysis, and Advances in the Era of Big Data,” *Information Sciences* 622 (2023): 178–210, <https://doi.org/10.1016/j.ins.2022.11.139>.
20. J. A. Donaghy, B. Jagadeesan, K. Goodburn, et al., “Relationship of Sanitizers, Disinfectants, and Cleaning Agents With Antimicrobial Resistance,” *Journal of Food Protection* 82, no. 5 (2019): 889–902, <https://doi.org/10.4315/0362-028X.JFP-18-373>.
21. S. Schubert, P. Darlu, O. Clermont, et al., “Role of Intraspecies Recombination in the Spread of Pathogenicity Islands Within the *Escherichia coli* Species,” *PLoS Pathogens* 5, no. 1 (2009): 1–10, <https://doi.org/10.1371/journal.ppat.1000257>.
22. M. A. Alhadlaq, O. I. Aljurayyad, A. Almansour, et al., “Overview of Pathogenic *Escherichia coli*, With a Focus on Shiga Toxin-Producing Serotypes, Global Outbreaks (1982–2024) and Food Safety Criteria,” *Gut Pathogens* 16, no. 1 (2024): 57, <https://doi.org/10.1186/s13099-024-00641-9>.
23. M. Ferdous, A. W. Friedrich, H. Grundmann, et al., “Molecular Characterization and Phylogeny of Shiga Toxin-Producing *Escherichia coli* Isolates Obtained From Two Dutch Regions Using Whole Genome Sequencing,” *Clinical Microbiology and Infection* 22, no. 7 (2016): 642.e1–642.e9, <https://doi.org/10.1016/j.cmi.2016.03.028>.
24. S. Shakil, H. M. Ali, R. Zarrilli, and A. U. Khan, “Extended Spectrum Beta Lactamases: A Critical Update,” *Multidrug Resistance A Global Concern* 18, no. 4 (2012): 115–129, <https://doi.org/10.2174/978160805292911201010115>.
25. I. F. Jaja, J. Oguttu, C.-J. I. Jaja, and E. Green, “Prevalence and Distribution of Antimicrobial Resistance Determinants of *Escherichia coli* Isolates Obtained From Meat in South Africa,” *PLoS ONE* 15, no. 5 (2020): e0216914, <https://doi.org/10.1371/journal.pone.0216914>.
26. National Department of Health, *Surveillance for Antimicrobial Resistance and Consumption of Antimicrobials in South Africa, 2021* (National Department of Health, 2021).
27. E. Taviani, A. Muchongo, S. W. Kim, J. A. S. Van Kessel, and B. J. Haley, “Genomic Analysis of Antibiotic-Resistant and -Susceptible *Escherichia coli* Isolated From Bovine Sources in Maputo, Mozambique,” *Foodborne Pathogens and Disease* 18, no. 6 (2021): 426–435, <https://doi.org/10.1089/fpd.2020.2901>.
28. F. Baquero, J. L. Martínez, V. F. Lanza, et al., “Evolutionary Pathways and Trajectories in Antibiotic Resistance,” *Clinical Microbiology Reviews* 34, no. 4 (2021): e0005019, <https://doi.org/10.1128/CMR.00050-19>.
29. T. J. Johnson and L. K. Nolan, “Pathogenomics of the Virulence Plasmids of *Escherichia coli*,” *Microbiology and Molecular Biology Reviews* 73, no. 4 (2009): 750–774, <https://doi.org/10.1128/mmbr.00015-09>.
30. A. Carattoli, “Resistance Plasmid Families in Enterobacteriaceae,” *Antimicrobial Agents and Chemotherapy* 53, no. 6 (2009): 2227–2238, <https://doi.org/10.1128/AAC.01707-08>.
31. Y. Xu, Y. Jing, L. Hu, et al., “IncFIB-4.1 and IncFIB-4.2 Single-Replicon Plasmids: Small Backbones With Large Accessory Regions,” *Infection and Drug Resistance* 2022 15 (2022): 1191–1203.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.

Supporting File 1. puh270098-sup-0001-SuppMat.docx