Plasmid 135 (2026) 102771

Contents lists available at ScienceDirect

Plasmid

journal homepage: www.elsevier.com/locate/yplas

Original Research Article

Comparative genomics of diverse *Escherichia coli* O157:H7 strains to characterize plasmids, prophages, virulence and antimicrobial resistance genes

Sydney Menzeko Gambushe [a], Peter Ayodeji Idowu [b,*], Oliver Tendayi Zishiri [a]

[a] Discipline of Genetics, School of Life Sciences, College of Agriculture, Engineering and Sciences, University of KwaZulu-Natal, Durban 4000, South Africa
[b] Section of Veterinary and Public Health, Department of Paraclinical Sciences, Faculty of Veterinary Science, University of Pretoria, South Africa

ARTICLE INFO

Keywords:
Escherichia coli
O157:H7
Plasmid
Virulence
Resistance
Prophage

ABSTRACT

Plasmids play a critical role in bacterial evolution and represent major drivers of the emergence and dissemination of antimicrobial resistance. As primary mobile genetic elements (MGEs), plasmids facilitate the horizontal transfer of resistance determinants alongside genes associated with virulence, metabolic functions, and broader adaptive advantages. Recent studies have further highlighted the importance of conjugative plasmids, such as IncI1-like elements, in mediating the spread of extended-spectrum β-lactamase (ESBL) genes and other clinically relevant traits across diverse bacterial populations. Whether the recurrent detection of these plasmids is coincidental or reflects unique genetic features that enhance their capacity for transmission remains an important question in microbial genomics. In this context, the present study analyses complete genome sequences and whole-genome maps of *Escherichia coli* O157:H7 strains to characterize their antimicrobial resistance genes, virulence-associated loci, prophage content, and plasmid profiles. Publicly available sequences from the NCBI GenBank repository were examined using comparative genomic tools, including BRIG, VirulenceFinder, ResFinder, PlasmidFinder, and PHASTEST. This work also underscores the limited availability of whole-genome data for *E. coli* O157:H7 and O157:H7NM in developing regions, particularly within African countries, highlighting the need for expanded genomic surveillance. Comparative analyses revealed that most strains displayed high genomic similarity to the reference Sakai strain, with relatively few missing regions, although a subset exhibited reduced homology marked by numerous gaps. Prophages, bacteriophages integrated into the bacterial genome, were found to contribute substantially to genomic diversity, influencing virulence potential, antimicrobial resistance, and patterns of horizontal gene transfer. These findings emphasize the complex role of mobile genetic elements in shaping the evolution of *E. coli* O157:H7 and reinforce the importance of continued genomic sequencing to further elucidate the pathogen's diversity and adaptive mechanisms.

1. Introduction

Escherichia coli (*E. coli*) is a significant cause of diarrheal disease and mortality worldwide, especially among children in developing countries (Kotloff et al., 2013; Lozano et al., 2012). In any case, a subset of *E. coli* is responsible for human diseases and has been linked to foodborne illnesses that have become more common over the last 20 years. *E. coli* O157 is the most common part of a group of pathogenic *E. coli* strains known as enterohaemorrhagic, verocytotoxin, or Shiga-toxin-producing organisms (Chapman et al., 1997; Abongo and Momba, 2009). The primary outbreaks linked to *E. coli* O157 occurred in Oregon and Michigan, USA, in, 1982, when it was identified in individuals who suffered from severe diarrhea and intense stomach pain after consuming cheeseburgers at a restaurant chain (CDC, 1982). Contamination can cause gastroenteritis that will be complicated by hemorrhagic colitis or hemolytic-uremic syndrome (HUS), which is the most common cause of acute renal failure in children. Shiga toxin (*Stx*)-producing *E. coli* (STEC) strains causing human diseases have a place in a huge, still-increasing number of O:H serotypes. Most episodes and scattered cases of hemorrhagic colitis and HUS have been ascribed to the STEC O157 strains (Rahimi et al., 2012; Tarr et al., 2005; Lockary et al., 2007; Werber et al., 2007).

In 2010, diarrheal diseases caused an estimated 1.4 million deaths worldwide, including more than 120,000 deaths related to

* Corresponding author.
E-mail address: ayodejiidowuolu@gmail.com (P.A. Idowu).

https://doi.org/10.1016/j.plasmid.2025.102771
Received 22 July 2025; Received in revised form 1 December 2025; Accepted 7 December 2025
Available online 8 December 2025
0147-619X/© 2025 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

enterotoxigenic *E. coli* (ETEC) and more than 88,000 deaths related to enteropathogenic *E. coli* (EPEC) (Lozano et al., 2012). Both ETEC and EPEC cause significant diarrheal morbidity and mortality in children. It occurs in developing countries (Kotloff et al., 2013; Nisa et al., 2013; Ochoa and Contreras, 2011). ETEC has also been identified as the leading cause of traveller's diarrhea in adults worldwide (Kotloff et al., 2013; Lozano et al., 2012; Qadri et al., 2005). Shiga toxin-producing *E. coli* (STEC) comprises a collection of genetically and phenotypically unique strains that exhibit significant variability in their ability to cause disease. Certain strains can lead to severe illness, whereas others may result in only mild diarrhea or no symptoms (Kotloff et al., 2013). This variability is partly attributed to their virulence gene repertoire. The genome of STEC is highly heterogeneous and is made up of a core genome that is preserved across STEC strains, along with accessory genomes that harbor genes specific to certain strains, lineages, or lesions (Lozano et al., 2012; Nisa et al., 2013; Ochoa and Contreras, 2011). Enterohemorrhagic *E. coli* (EHEC) was initially defined as STEC strains responsible for causing hemolytic uremic syndrome (HUS) or hemorrhagic colitis (HC). It produces enteric hemolysin and Shiga toxin (*Stx*), leading to attachment loss syndrome associated with (A/E) intestinal epithelial cell lesions (Qadri et al., 2005). Although the EHEC prototype, STEC O157:H7, is recognized as the most prevalent cause of outbreaks associated with STEC, non-O157 strains are increasingly identified as EHEC because of their connection to HUS. Certain strains, such as O157:H7, are representative of typical EHEC strains. (Nataro and Kaper, 1998; Kaper et al., 2004).

Foodborne pathogenic *E. coli* O157:H7 is responsible for over 96,000 cases of diarrheal illnesses and approximately 3200 hospitalizations in the United States annually (Scallan et al., 2011). This pathogen can result in severe gastrointestinal issues, including intense diarrhea, which may escalate to more severe complications such as hemolytic uremic syndrome (HUS) and even lead to death (Karmali et al., 2010). The *eae* (which encodes intimin) and *Stx* (which encodes Shiga toxin) found in foodborne pathogenic *E. coli* O157:H7 strains are crucial to the development of HUS (Paton and Paton, 1998). Furthermore, a study that was done by Robinson et al. (2006) showed that the Shiga toxin produced by *E. coli* O157:H7 can facilitate attachment to epithelial cells and promote colonization in the intestines of mice. In contrast to *Stx*-producing *E. coli* O157:H7, strains that do not produce *Stx* do not generate Shiga toxin. While these strains can cause symptoms such as diarrhea, they are generally not associated with HUS, despite carrying virulence factors like the *eae* and *bfpA* genes (Ochoa and Contreras, 2011; Ferdous et al., 2015).

Diarrheagenic *E. coli* is currently categorized into a limited number of groups, primarily based on the presence of specific virulence genes, and is classified under a single pathogenic variant (pathovar). EPEC strains that possess the Locus of Enterocyte Effacement (LEE) region and Bundle-Forming Pilus (BFP) are referred to as typical EPEC (tEPEC), while those that lack the BFP are designated as atypical EPEC (aEPEC) (Nisa et al., 2013; Nataro and Kaper, 1998; Kaper et al., 2004). In the meantime, ETEC is identified at the molecular level by the presence of heat-labile (LT) or heat-stable (ST) enterotoxins, along with a variety of accessory virulence factors, including the *eae* autotransporter (Fleckenstein and Kuhlmann, 2019; Fleckenstein et al., 2010). These canonical destructiveness variables are frequently encoded on plasmids or other mobile components within the segregates from each of these pathovars (Kaper et al., 2004; Croxen and Finlay, 2013). Regardless, through the identification of canonical virulent traits, several clinical isolates can be organized into various pathovars.

With the sequencing of additional strains of the same species, new characteristics are consistently uncovered. An examination of more than 2000 *E. coli* genomes showed an extensive pangenome (approximately 89,000 genes in total) and a limited core genome (around 3100 genes, which are maintained across all genomes) (Croxen and Finlay, 2013), which is attributed at least in part to its adaptable lifestyle (Anjum et al., 2018; Brockhurst et al., 2019). STEC alternates between host and non-host conditions. The additional genome can be viewed as a reserve for STEC, equipping them effectively to adjust to different settings: whether from herbivorous mammals, interaction with the surroundings, or interaction with people. Indeed, the accessory genome is believed to contain information specific to particular niches and can help in tracing the origins of isolates. (Lupolova et al., 2019; Barth et al., 2016; Arimizu et al., 2019; Carter, 2017). The STEC comparative genomics showed a similar developmental trajectory: the gaining and losing of virulence genes occurred simultaneously at different points in distinct lineages (Nataro and Kaper, 1998; Kaper et al., 2004; Grad et al., 2013). The significance of Horizontal Gene Transfer (HGT) in the evolution of the genome and pathogenicity of STEC is far more comprehensive than we previously realized. The EHEC strain EDL933 and the *E. coli* K-12 strain MG1655 comparative genomic study revealed more than one thousand novel genes in EDL933 (Perna et al., 2001). These genes are typically arranged into clusters of varying sizes referred to as genomic islands (GIs) or islands of pathogenicity (PAIs) when virulence genes are present. GIs offer unique benefits to an organism and improve the capabilities of bacteria; a portion of this arises from the genes that come with a new GI. (Dobrindt et al., 2004). In the context of STEC, aside from harmful traits, characteristics associated with resistance to stress, self-aggregation, adherence, and biofilm formation are linked to genomic islands (GIs) (Montero et al., 2017; Mercer et al., 2017). These GIs are usually sizable, varying between 10 and 200 Kb, and they insert themselves into the chromosome close to tRNA genes along with neighbouring direct repeats (DRs), whereas transposons or insertion sequences (Iss) are frequently found near their termini. (Juhas, 2019). Consequently, GIs signify instability segments of the genetic material that can undergo deletion, rearrangement, and transfer, which play a role in the rapid evolution of STEC. (Montero et al., 2017; Bielaszewska et al., 2007).

The conventional methods that distinguish between *Stx*-positive and *Stx*-negative *E. coli* O157:H7 are constrained by their emphasis on particular selected DNA sequences in the bacterial genome, such as specific genes and/or short DNA segments. With advancements in next-generation sequencing (NGS) technologies and whole-genome sequencing (WGS), scientists can now evaluate and compare DNA sequences throughout the bacterial genome. Antimicrobial resistance (AMR) poses a global public health challenge and has multifaceted effects on the food production sector. The extensive use of antimicrobials in livestock production is widely acknowledged as a primary factor contributing to antimicrobial resistance in both people and animals. Antimicrobials are utilized for purposes including therapeutic and non-therapeutic usage, such as metaphylactic, prophylactic, and development promoters (Ochman et al., 2000). Subsequently, the development and spread of AMR over the farm-to-plate continuum put occupationally uncovered specialists (viz. ranchers, rural specialists, abattoir specialists, nourishment handlers), their near contacts and customers after the nourishment chain at hazard of defilement or contamination by antimicrobial safe microscopic organisms and/or antimicrobial resistance genes (ARGs) (Ochman et al., 2000; Escobar-Páramo et al., 2004). ABR anticipation and control measures centre not only on people but also on creatures and their related situations (Escobar-Páramo et al., 2004).

Acquiring factors that lead to destruction via the evolving standard of quality exchange may pose a significant limitation that fuels the development and expansion of harmful microorganisms, in contrast to the modification of current DNA (O'Neill, 2016). A particular genomic foundation may be required for the integration, maintenance, and expression of remote DNA (O'Neill, 2016; Founou et al., 2018), and the advancement of pathogenic microbes regularly shows a solid ancestry reliance. Strains with the same pathotype have often developed from numerous heredities, but the hereditary factors behind such HGT are not completely identified. Genotypic approaches have been proposed as a substantial alternative to phenotypic AST since the early 1990s (Fleckenstein et al., 2010; Courvalin, 1994).Although genotypic tests have been and are broadly utilized for the discovery of AMR qualities,

their constrained sensitivity has prevented their application as reference AST strategies. (Croxen and Finlay, 2013; Anjum et al., 2018; Brockhurst et al., 2019). With increased accessibility and reduced costs, whole-genome sequencing (WGS) and next-generation sequencing (NGS) have emerged as a viable method for the routine characterization of microorganisms. NGS methods surpass other genotypic techniques for antimicrobial resistance (AMR) detection because they enable the identification of any recognized AMR gene or mutation and the discovery of new variants of established AMR factors (Anjum et al., 2018; Brockhurst et al., 2019). Furthermore, sequencing data can be stored indefinitely and can be re-evaluated as new AMR determinants are discovered through phenotypic methods (Anjum et al., 2018; Brockhurst et al., 2019). The primary limitation associated with any genotypic antimicrobial susceptibility testing (AST) approach is represented by the fact that only known AMR mechanisms can be detected, while resistance arising from novel mechanisms and/or alterations in genes, such as increased efflux pump activity or hetero-resistance, may be overlooked. Over the last ten years, at least forty-seven open-access bioinformatics pipelines have been developed (Anjum et al., 2018; Brockhurst et al., 2019).

For identifying AMR qualities in NGS, information has been distributed in more than 100 countries (surveyed by Hendriksen et al., 2019; Bortolaia et al., 2020; Lupolova et al., 2019) and has been created, but they generally need simple and quickly interpretable yields, including interpretation of genotypes into anticipated phenotypes (Bortolaia et al., 2020; Hendriksen et al., 2019). ResFinder was extended with PointFinder, a platform that recognizes chromosomal point transformations intervening resistance to selected antimicrobial agents in several chosen bacterial species (Bortolaia et al., 2020; Zankari et al., 2017). Whole genome sequences accessible in STEC have shown significant qualities due to the level of HGT and genomic changes (Bortolaia et al., 2020; Acar et al., 1970; Davies et al., 2020; Mouton et al., 2018; Bader et al., 2018). Employing comparative genomics, the identification of pathogenicity and resistance traits, along with monitoring harmful effects, can be accomplished through the associated plasmids of microorganisms that present a threat to community health. In the current investigation, the primary focus is to analyze the complete genome sequences from available *E. coli* O157:H7 strains identified in NCBI GenBank to explore possible resistance, virulence, and plasmid traits to distinguish between the strains. A full genome mapping for comparative genomics was conducted on the *E. coli* O157:H7. (Bortolaia et al., 2020; Davies et al., 2020; Mouton et al., 2018).

The global genome structure accessible by STEC exhibits widely varying quality due to horizontal gene transfer and genomic variation (Bortolaia et al., 2020; Acar et al., 1970; Davies et al., 2020; Mouton et al., 2018; Bader et al., 2018). Comparative genomics distinguishes evidence of harm from safety, distinguishes between related plasmids, and detects pathogenic microorganisms that pose clear health risks. The main objective of the investigation was to identify potential resistance genes, virulence genes, prophages, and plasmid properties among selected strains of *E. coli* O157:H7 from accessible strains in the GenBank repository. Analyze the complete genome sequences and whole-genome mapping to investigate the genomic differences among the strains of O157:H7 (Bortolaia et al., 2020; Acar et al., 1970; Davies et al., 2020; Mouton et al., 2018; Bader et al., 2018). Limited studies of the ecology of *E. coli* O157 have been reported, particularly from developing countries, more especially in African countries, as there is limited data availability in terms of Whole Genome Sequences of *E. coli* O157:H7 and *E. coli* O157:H7NM. Therefore, this study illustrates, using the available data in the NCBI GenBank system, the importance of conducting more genome sequences of this bacterium to study its pathogenicity and antimicrobial resistance.

2. Materials and methods

2.1. National center for biotechnology information (NCBI)

Complete genome sequences of *E. coli* O157:H7 were retrieved from the NCBI Taxonomy Browser and associated GenBank records. An initial search for *E. coli* O157:H7 identified 136 available sequences. For this study, only fully assembled, complete genome sequences were selected, while laboratory strains and incomplete assemblies were omitted. Strains with insufficient metadata or lacking usable sequence information were omitted to ensure consistency in comparative genomic analyses.

2.2. BLAST ring image generator (BRIG) analysis

Comparative chromosomal maps were generated using the BLAST Ring Image Generator (BRIG), which constructs circular visualizations based on BLAST alignments (Alikhan et al., 2011). The annotated chromosome of the *E. coli* O157:H7 Sakai strain (RIMD 0509952) served as the reference for all comparisons. Each selected genome was aligned against the reference to assess sequence conservation, regions of divergence, and the presence or absence of genomic islands. All analyses were performed using the default BRIG parameters unless otherwise specified.

2.3. Plasmid identification

The PlasmidFinder database (https://cge.cbs.dtu.dk/services/PlasmidFinder/) was utilized to identify the presence of plasmids within the *E. coli* O157:H7 strains (Carattoli et al., 2014; Camacho et al., 2009). Every sequence used in this study was consolidated into a single document and then transferred to the PlasmidFinder platform. The platform provides four selection options: database selection, minimum % identity threshold, read type selection, and minimum % coverage threshold. Default settings were applied, with the minimum % identity and coverage thresholds set to 95 % and 60 %, respectively. The Enterobacteria database was chosen, and the analysis was performed using grouped or draft genomes/contigs as the selected read type.

2.4. Resistance identification

The ResFinder database (https://cge.cbs.dtu.dk/services/ResFinder/) was utilized to identify antimicrobial resistance genes (Camacho et al., 2009; Bortolaia et al., 2020; Zankari et al., 2017). All relevant sequences were combined into one file and uploaded to the platform. The tool offers four selection options: chromosomal point mutations, acquired antimicrobial resistance genes, species selection, and read type selection. For this analysis, chromosomal point mutations and acquired antimicrobial resistance genes were chosen. *E. coli* was selected as the species and assembled, or draft genomes/contigs were designated as the read type.

2.5. Virulence gene identification

The VirulenceFinder database (https://cge.cbs.dtu.dk/services/VirulenceFinder/) was employed to identify virulence genes (Camacho et al., 2009; Joensen et al., 2014; Malberg Tetzschner et al., 2020). All relevant sequences were combined into one file and uploaded to the platform. The tool provides four options for selection: species choice, identity percentage threshold, minimum length, and type of read. The default values for the identity percentage threshold and minimum length were used, set at 90 % and 60 %, respectively. *E. coli* was chosen as the species, and assembled, or draft genomes/contigs, were indicated as the type of read. Then, the double-hierarchical clustering heatmap, which shows the resemblance of the strains, was generated using Python programming language (v3.11), and the pairwise distance calculations among strains were determined using the Jaccard distance metric.

2.6. PHASTEST

The identification of prophages within the chromosomes of all strains was accomplished by retrieving the complete chromosomal sequences of each strain from FASTA files provided by NCBI and subsequently uploading them to PHASTEST, as PHASTER was undergoing maintenance at that time (Zhou et al., 2011; Arndt et al., 2016). Prophages were categorized into three types: intact, questionable, and incomplete, based on their scores. Scores exceeding ninety are categorized as intact, scores between 70 and 90 as questionable, and scores below 70 as incomplete.

2.7. Comparative genomic visualization and data extraction

Whole-genome analyses were conducted using the BLAST Ring Image Generator (BRIG) (Alikhan et al., 2011) to visualize the sequence similarity between the *E. coli* O157:H7 Sakai reference genome and various query genomes obtained from NCBI GenBank. Each BRIG map illustrates the reference genome at the center, encircled by concentric rings that represent query strains.

Sequence identity was calculated using BLASTn with default settings, and color gradients reflect percentage identity throughout the genome (ranging from 70 % to 100 %). To facilitate comparative assessments, data from the BRIG-generated circular alignment maps were extracted and transformed into categorical matrices. The presence or absence of homologous regions, along with the observations of conserved or divergent genomic segments (such as prophage regions, virulence islands, and plasmid-associated loci), were visually inspected and documented from the BRIG output images. This visual information was then converted into binary format (1 = present, 0 = absent) to depict sequence conservation across each strain in comparison to the reference genome.

Furthermore, PlasmidFinder results were similarly processed for plasmid replicon analysis. The output images, color-coded, were analyzed, and the presence/absence data of replicons were transcribed into a binary matrix corresponding to the various *E. coli* O157:H7 strains. This dataset was utilized to create a double-hierarchical clustered heatmap, illustrating similarity patterns among strains and plasmid types. Data visualization and clustering were executed in Python (v3.12) using the pandas, NumPy, matplotlib, and seaborn libraries. Clustering was performed using the Euclidean distance metric and average linkage method to organize strains and attributes based on shared presence/absence patterns (supplementary information). Furthermore, a heatmap was generated using Heatmapper heatmapper. ca/expression/ Clustering was performed using the Euclidean distance matrix and average linkage method.

3. Results

3.1. BRIG

Whole-genome analyses performed using BRIG, together with average nucleotide identity (ANI) assessments, revealed that *E. coli* O157:H7 strains exhibit substantial genomic conservation relative to the Sakai reference strain. Most isolates demonstrated greater than 90 % genome coverage and high ANI values, consistent with the presence of a strongly conserved core genome characteristic of this lineage (Fig. 1; Table 1). A subset of strains shows reduced homology with ANI and identity values that drop into the low 90s or lower, corresponding to genomes with multiple gaps and divergent accessory regions. Divergent loci are concentrated near 1.5–2.5 Mb and 4.5–5.0 Mb in regions enriched in prophage insertions and mobile elements. These observations indicate a conserved core genome with variable accessory regions contributing to observed genomic diversity. Nonetheless, several strains, including EC4115, EC4501, EDL933, EC4042, and EC4045, displayed near-complete alignment with the reference genome, reflecting minimal

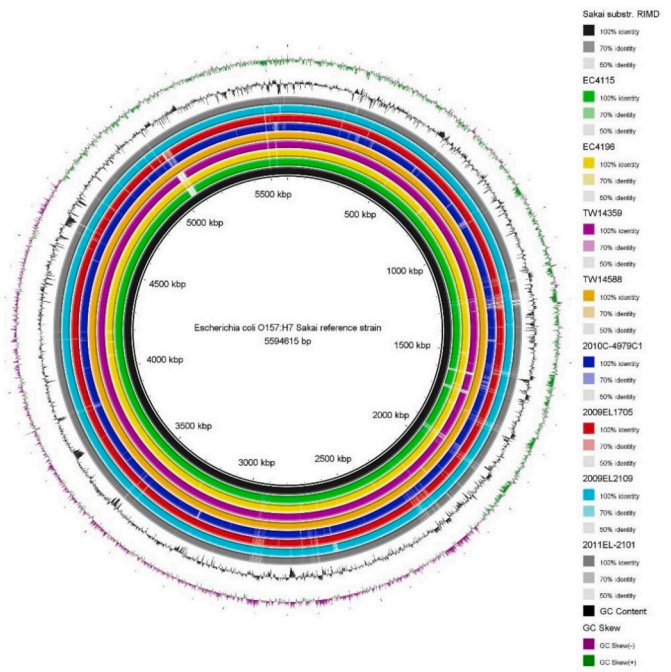

Fig. 1. A representative figure of comparison of BLAST Ring Image (BRIG) generated for *E. coli* O157:H7 strains, illustrating genomic similarities and differences between a reference strain and multiple query strains. The analysis was conducted using data obtained from NCBI, focusing on diverse strains (EC4115, EC4196, TW14389, TW14588, 2010C-4979C1, 2009EL1705, 2009EL2109, 2011EL-2101). The BLAST Ring Image highlights the genomic alignments, enabling a comparative overview of genetic variations and conserved regions across the selected strains.

genomic divergence.

Although the overall genome structure was highly conserved, BRIG mapping also identified a gradient of variability across the collection of isolates. A subset of strains exhibited moderate to pronounced divergence, with ANI values ranging from approximately 80 % to 96 % and genome coverage between 94 % and 98 % in the most variable genomes. These differences were predominantly concentrated within genomic regions known to contain prophages, plasmid-associated sequences, and the locus of enterocyte effacement (LEE). Divergent zones were consistently observed between approximately 1.5–2.5 Mb and 4.5–5.0 Mb, corresponding to areas enriched in mobile genetic elements and recognized hotspots for recombination and horizontal gene transfer. Such patterns align with previously described genomic plasticity within *E. coli* O157:H7 (Ogura et al., 2015; Eppinger et al., 2011).

Across the dataset, most isolates retained key virulence determinants associated with pathogenicity, including Shiga toxin genes (*Stx*) and the intimin gene (*eae*). However, several divergent strains displayed reductions or gaps within plasmid-associated regions, particularly those encoding genes such as *ehxA* and *katP*, suggesting variable retention of accessory virulence factors. These gene losses corresponded with strains exhibiting lower homology to the Sakai reference and were typically associated with broader genomic deletions or rearrangements affecting mobile elements.

Collectively, the BRIG analysis demonstrates that while the core genome of *E. coli* O157:H7 is highly conserved across strains, notable differences arise primarily within regions enriched for prophages, plasmid sequences, and other mobile genetic elements. These variable loci contribute to the structural and functional diversity observed among isolates and reflect the dynamic evolutionary processes shaping the genomic architecture of this pathogen.

Table 1
Comparative Genomic Similarity Between *E. coli* O157:H7 Sakai Reference Strain and Other Strains Based on BRIG Analysis.

STRAIN ID	SOURCE	GENOME SIZE (MB)	COVERAGE VS SAKAI (%)	AVERAGE IDENTITY (%)	UNIQUE / DIVERGENT REGIONS	NOTABLE GENETIC DIFFERENCES REMARKS
RIMD 0509952 (SAKAI)	Human (Reference)	5.59	100	100	—	Reference genome; baseline for all comparisons
EC4115	Human	5.56	99.5	99–100	Few prophage regions	Minor phage island differences, highly conserved strain
EC4196	Bovine Reservoir - Human intestinal microflora	5.54	99.2	98–100	Divergence near 4.5 Mb	Small plasmid variation, nearly identical to Sakai
TW14359	Spinach bag – Human intestinal microflora	5.53	98.8	97–99	Small deletions in prophage regions	Loss of certain phage-encoded loci, moderate variation
TW14588	Taco John Outbreak – lettuce - Human	5.52	98.5	96–98	Several prophage islands are absent	Reduced prophage content; plasmid loss, distinct variant
EC4076	Human	5.54	99.6	99.8	Minor phage variation	Minimal deletions, nearly identical to the reference
EC4113	Spinach bag – Human intestinal microflora	5.53	99.4	99.7	Gaps in plasmid genes	Rearrangements in pO157, highly conserved
EC4009	Human	5.6	>99.5	~98–99	The majority of the genome is conserved with few scattered low-identity zones	High similarity
EC4191	Spinach bag – Human intestinal microflora	5.6	99.0	96–98	Large, conserved regions with low identity spots	Slight divergence; nearly identical to the reference
EC4205	Bovine Reservoir - Human Intestinal microflora	5.6	98.5–99.0	95–96	Generally conserved with mild/ moderate gaps	A few divergent loci due to phage or plasmid differences
EC4084	Human	5.6	98.5–99	95–97	Most regions are conserved with moderate variation in 3–5 %	Regions of lower identity suggest divergence in mobile or phage-associated genes.
EC4192	Human	5.6	99	96–98	Highly conserved	Similar profile to EC4009, but with distinct small-variable regions.
EC4127	Human	5.6	>99.5	98–99	Very similar to the Sakai strain with a few minor gaps	Very closely related; minimal divergence
FRIK2000	Bovine Reservoir - Human intestinal microflora	5.5–5.6	98.5–99	95	Most of the genome is conserved, with notable unique segments	Differences may involve virulence islands or horizontally transferred elements.
EC4486	Human	5.50	99.0	99.5	Minor LEE deletions	Possible impact on A/E lesion genes, stable but slightly varied
EC4501	Taco John Outbreak – Human intestinal microflora	5.55	99.5	99.8	Minimal phage variation	Highly conserved, very high identity
EC508	Human	5.53	99.3	99.7	Small prophage deletions	Loss of minor islands at 4.5 Mb, slightly reduced content
EC536	Human	5.6	98–98.5	93–95	Moderate to high conservation with several moderate gaps	Genetically similar, but more variable than EC4127
FRIK966	Bovine Reservoir - Human intestinal microflora	5.50	99.0	99.4	Missing plasmid fragments	Reduced virulence plasmid, slight plasmid loss
G5101	Human	5.56	97	93–95	Majority conserved with noticeable gaps at 1.2 Mb, 4.3 Mb	Contains more variation; potential loss/acquisition of pathogenicity islands
EDL933	Ground hamburger outbreak - Human	5.55	99.6	99.8	Minor prophage deletions	Nearly identical to Sakai, classical outbreak strain
F8092B	Human	5.53	99.4	99.6	Variation at phage loci	Retains key virulence genes, stable clinical strain
T1543_06	Cattle	5.52	99.2	99.5	Deletions near 1.8 Mb	Phage rearrangements, slight divergence
SS52	Environment/food	5.56	99.8	99.6	High identity; stable genome structure	Broad conservation, minor prophage loss
TW14313	Human	5.54	99.7	99.3	Minor variation due to mobile elements	Conserved backbone, limited deletions near phage regions,
G5303	Missing	5.58	99–100	98–99	Closely related to the reference strain	Highly conserved, minor gaps at 3.8 Mb, 5.2 Mb
H093800014	Not disclosed	5.58	99–100	97–98	Likely outbreak-associated	Conserved, divergence at 2.5 Mb, 4.5 Mb
EC4042	Human	5.58	100	98–99	Nearly identical; Closely related to the reference strain	Entire genome, few gaps at 2.5 Mb, and 3.5 Mb
EC4045	Human	5.58	100	98–99	Genetically similar to EC4042	Entire genome, minimal variability
EC4206	Bovine Reservoir - Human intestinal microflora	5.57	98	95–97	Possible horizontal gene transfer	Large, conserved regions, unique loci at 1 Mb, and 4 Mb
1044	Human	5.58	99–100	98–99	Nearly identical to Sakai	Highly conserved, small gaps near 3 Mb, 5.5 Mb
1125	Human	5.58	99–100	97–98	Minimal gene acquisition	Highly conserved, few variable loci (4.7 Mb)

(*continued on next page*)

Table 1 (*continued*)

EC1212	Human	5.57	98–99	95–96	Mobile genetic element variation	Large, conserved core, variable segments (1 Mb, 4 Mb)
LSU-61	Human	5.56	97–98	93–95	Local evolution; associated with outbreak variant	Core conserved, gaps at 2 Mb, 3.5 Mb
SS17	Bovine RSE cells	5.57	98–99	95–97	Evidence of recombination	Conserved backbone, gaps near 1.5 Mb, 4.8 Mb
EcK2188	Missing	5.53	99.5	99.0	Core genes preserved with variation in prophage and plasmid loci	Typical environmental variation pattern and adaptation
EcK4396	Missing	5.51	99.3	98.8	Rearranged mobile elements	Conserved, prophage gaps
EcK5453	Missing	5.50	99.1	98.5	Phage-related gene loss	Core conserved, deletions near 4.5 Mb
EcK5806	Missing	5.51	99.3	98.8	Possible host-specific adaptation	Conserved backbone, variable prophage regions
EcK7140	Missing	5.49	99.2	98.5	Reduction in phage/island content	Core conserved, gaps at 2–2.3 Mb
EcK6590	Missing	5.47	98.9	98.0	Most divergent; reduced Sakai-like prophages	Core retained, loss of virulence islands
08BKT61141_454	Cattle	5.49	98.8	98.2	Host-specific adaptation	Core genome intact, variation in phage loci
09BKT048303_4541	Cattle	5.52	94	80–85	Fragmented conserved regions with several unique regions at 2.7 Mb and 4.5 Mb	Most divergent from the reference strain
06–4039	Human	5.54	95–96	88–91	Conserved islands are scattered with gaps distributed throughout	Shows moderate divergence; GC skew indicates HGT events.
07–3091	Human	5.55	96–97	91–93	The majority of the backbone is conserved, with unique regions at 2.5 Mb and 4.8 Mb	More variation than human strains; possibly environmentally acquired genes.
07–3391	Human	5.54	95	85–90	Core conserved, fragmented regions with distinct, unique regions at 3.2 Mb	High variability; likely adapted to a non-human host or environment.
08–3037	Human	5.4–5.5	96.5–97.5	88–90	Patchy conservation with many gaps; low similarity segments	Most divergent strain in this group; distinct evolution, possibly environmental or zoonotic
08–3527	Human	5.55	96	Highly divergent	Highly divergent to the reference strain	Moderate conservation, unique regions at 2.8, 4.6 Mb
08–4169	Human	5.54	96	88–91	Reduced similarity; niche adaptation	Scattered conserved islands, many gaps
2011EL-2099	Missing	5.53	95	85–90	Highly divergent to the reference strain	Some conserved backbone, large divergent regions
06–3745	Missing	5.52	94	80–85	Significant divergence; gene loss/gain	Core genome only, several large gaps (2–4 Mb)
2009EL1449	Missing	5.55	96–97	90–92	Highly core conserved to the reference strain	Conserved core, large gaps at ~1 Mb, 4.2 Mb
2009EL1913	Missing	5.54	95–96	88–90	Phage acquisition/loss is evident	Scattered conserved regions, gaps at 2–4 Mb
08–4529	Human	5.53	94–95	85–88	Highly divergent to the reference strain	Core moderately conserved, large gaps
2009C-42,581	Missing	5.52	94	80–85	Reduced virulence potential	Fragmented core, large genome-wide gaps
2010C-4979C1	Missing	5.51	98.2	95–98	Gaps at 1.5–2.0 & 4.8–5.0 Mb	Possible adaptation to a non-human host, slightly divergent
2009EL1705	Missing	5.50	97.5	94–97	Several low-identity regions	Missing katP, partial ehxA, reduced plasmid content
2009EL2109	Missing	5.49	97.0	93–96	Multiple gaps across the genome	Truncated LEE pathogenicity island, lower virulence potential
2011EL-2101	Missing	5.48	96.8	93–95	Multiple low-identity loci	Major deletions in plasmid genes (*ehxA*, *katP*), the most divergent in the group
2011EL-2103	Missing	5.49	98.5	98.9	Missing prophage elements	Gaps near tRNA-phage regions, environmental adaptation
2011EL-2104	Missing	5.48	98.2	98.7	Several missing regions	Loss in plasmid regions, slightly divergent
2011EL-2105	Missing	5.47	98.0	98.6	Plasmid/mobile element gaps	Mild divergence to the reference strain
2011EL-2106	Missing	5.46	97.8	98.4	Multiple genomic deletions	Large-scale divergence, lowest similarity in the set
2011EL-2107	Missing	5.49	98.5	98.9	Gaps in prophage sequences	Moderate divergence
2011EL-2108	Missing	5.47	98.3	98.8	Missing mobile elements	Variations at 1.5–2.0 Mb, phage region divergence
2011EL-1107	Missing	5.46	98.0	98.6	Deleted plasmid genes	Divergent plasmid content, lower virulence potential
2011EL-2090	Missing	5.45	97.7	98.4	Extensive genomic gaps	Loss of prophage/plasmid regions, most divergent isolate
2011EL-2091	Missing	5.5	98	93–95	The majority, with minor gaps and some localized divergence	Good conservation; possibly environmental or zoonotic source.
2011EL-2092	Missing	5.5	97.5–98.5	92–94	Substantial, but less than 2091, but broader scattered variations	Suggests divergence due to non-human host adaptation.
2011EL-2093	Missing	5.5	97–98	90–92	Moderate with increased variable content	More divergent, possibly due to host-specific adaptations.
2011EL-2094	Missing	5.5	96.5–97.5	88–90	Least conserved, with many unique or missing regions	Most divergent; many absent regions suggest loss/gain of genes or mobile elements.
2011EL-2096	Missing	5.5	98.0	93–95	Strong conserved backbone with moderate divergence in select regions	Similar to EC536 in pattern; likely zoonotic link
2011EL-2097	Missing	5.5	97.5–98.0	91–93	Conserved core with broader low-identity or missing areas	More divergence than 2096; unique gene regions visible
2011EL-2098	Missing	5.5	97.0–97.5	90–92	Reduced conserved regions with several variable islands	The pattern suggests adaptation or gene loss/acquisition

3.2. Plasmids

Analysis of plasmid replicons across the 76 *E. coli* O157:H7 genomes revealed the presence of eleven distinct plasmid types: *IncFIB (AP001918)*, *IncFII*, *IncFIA*, *pEC4115*, *IncI1-I (Alpha)*, *IncB/O/K/Z*, *Col156*, *IncI2 (Delta)*, *IncN*, *Col(BS152)*, and *IncFII (pSE11)* (Figs. 2–3). The distribution of these plasmid replicons varied widely among isolates, reflecting considerable diversity within the accessory genome of this lineage.

The most frequently identified plasmid families were *IncFIB* and *IncFII*, both commonly associated with virulence plasmids characteristic of *E. coli* O157:H7. Several replicons, including *IncB/O/K/Z*, *Col156*, *IncN*, *Col(BS152)*, and *IncFII (pSE11),* were detected only in a limited number of isolates, indicating that certain plasmid types are sporadically distributed across the dataset. These replicons often co-occurred with regions of genomic divergence involving mobile genetic elements, suggesting a potential association between plasmid carriage and variability in prophage or accessory gene content.

Overall, the heterogeneous pattern of plasmid replicon distribution is consistent with the known genomic plasticity of *E. coli* O157:H7. Prior studies have reported that plasmids carrying multiple replicons contribute to adaptability and may facilitate the transfer of virulence or antimicrobial resistance determinants within this serotype (Zhang et al., 2021a). The observed diversity underscores the dynamic role of plasmids in shaping strain-to-strain variation and accessory genome composition.

3.3. Virulence genes

A total of 53 virulence-associated genes were identified across all *E. coli* O157:H7 genomes, encompassing categories such as toxins, LEE-encoded and non-LEE-encoded type III secretion system (T3SS) effectors, adherence factors, secretion systems, iron acquisition mechanisms, and autotransporters (Figs. 2–3). Core virulence genes, including *aslA, chuA, espA, espB, fimH, tir, nleC, nlpI,* and *terC,* were consistently present in all isolates, underscoring their essential role in the pathogenic potential of this serotype.

Variation was observed in Shiga toxin (*Stx*) subtypes among the isolates. While *Stx1* and *Stx2* genes were broadly conserved, the specific variants displayed considerable diversity, with multiple *Stx2* subtype combinations detected across the dataset. This heterogeneity aligns with prior observations that *Stx*-encoding prophages are major contributors to genomic variability within *E. coli* O157:H7 (Karch et al., 2005).

Several accessory virulence genes also exhibited variable distribution patterns among strains. Genes such as *cba, efa1, colE8, ehxA,* and *fdeC* were present in some isolates but absent in others, reflecting natural variation within the accessory genome. These differences corresponded with regions of genomic divergence associated with plasmid

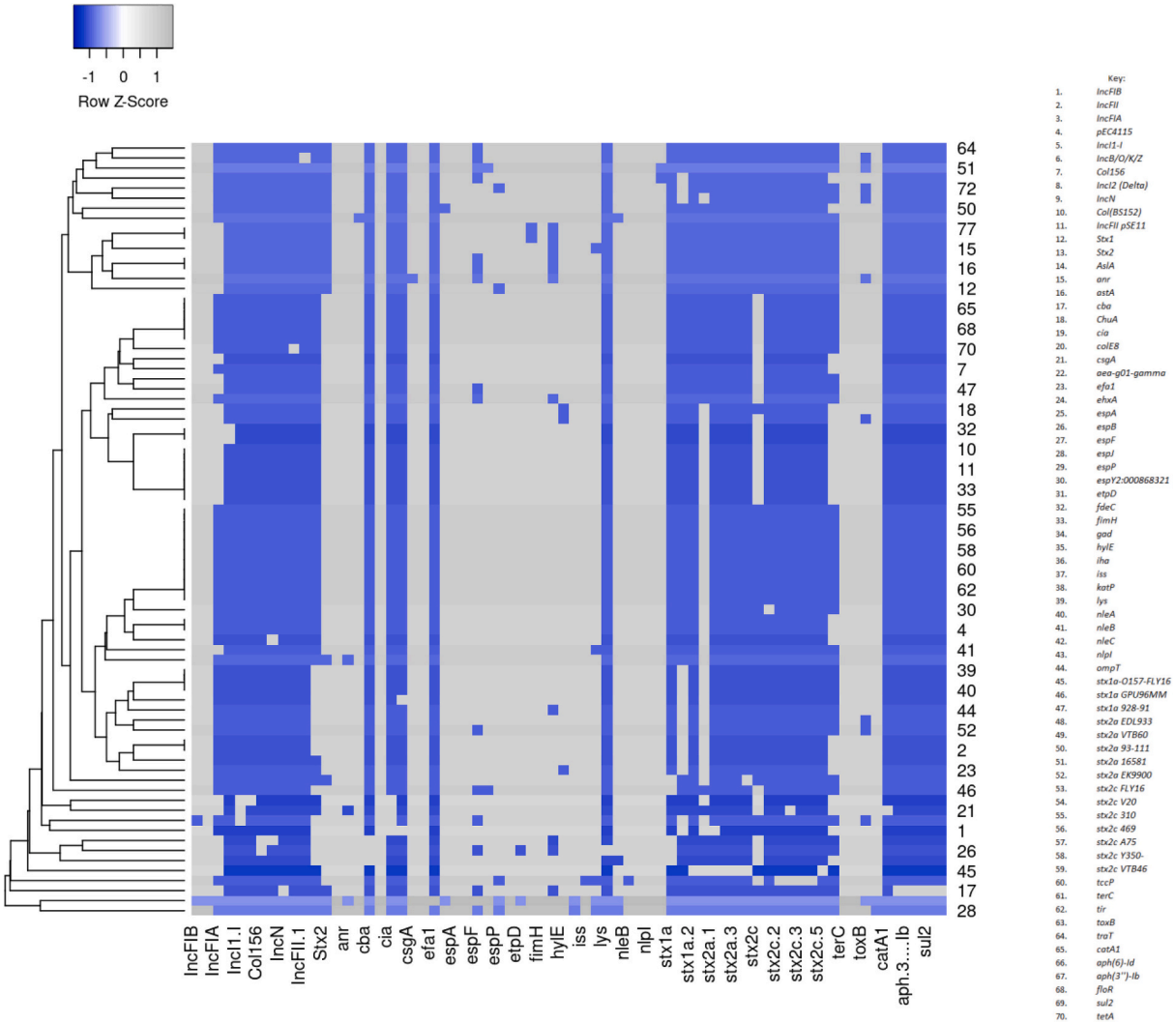

Fig. 2. A double-hierarchical clustering heatmap using Heatmapper software (clustering plasmids, virulence, and antimicrobial resistance genes) of *E. coli* O157:H7 strains versus genetic attributes. Parameters used include Jaccard distance with average linkage (1 = present = blue, 0 = absent = grey).

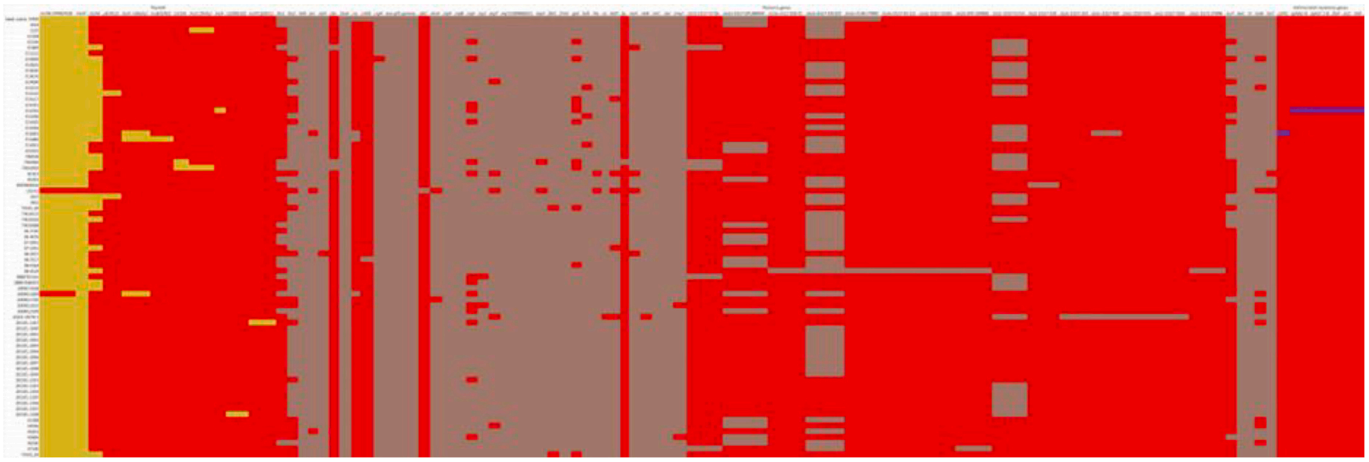

Fig. 3. Hierarchical clustering heatmap of plasmid replicons, virulence factors, and antimicrobial resistance genes among *Escherichia coli* O157:H7 isolates.

content and prophage insertions, suggesting that mobile genetic elements play a central role in shaping strain-specific virulence profiles.

Despite this diversity, LEE-encoded T3SS effectors remained highly conserved across all genomes analyzed, emphasizing their fundamental role in epithelial adherence and colonization (Wickham et al., 2019). Overall, the virulence gene analysis indicates that *E. coli* O157:H7 maintains a stable core virulence repertoire, with variation primarily confined to accessory loci associated with mobile genetic elements.

3.4. Prophage analysis (PHASTEST)

Prophage analysis using PHASTEST revealed that all 76 *Escherichia coli* O157:H7 genomes contained phage-like elements, consistent with the mosaic genomic architecture characteristic of this lineage (Herold et al., 2004). The Sakai reference genome harbored 22 prophage regions, including 13 intact, 5 questionable, and 4 incomplete elements, with a GC content of 49.86 % (Table 2).

Across the dataset, the number and composition of prophages varied substantially among isolates. Prophage counts ranged from as few as 6 to as many as 30 regions per genome, reflecting notable differences in the extent of mobile element acquisition or loss. The highest prophage loads were observed in isolates such as EC4113, EC869, and EC4501, whereas lower prophage numbers were recorded in isolates including EC4205, EC4009, and EC4084. Several divergent strains, such as 2010C-4979C1, K4396, and K6590, displayed prophage profiles distinct from those of more conserved genomes, indicating independent histories of phage integration and recombination (Tables 2–3).

These patterns underscore the central role of prophages in contributing to genomic diversification within *E. coli* O157:H7. Variation in prophage content corresponded closely with regions identified in BRIG analyses as sites of reduced sequence identity or structural variability, supporting previous findings that prophages act as major drivers of accessory genome evolution in this serotype (Brüssow et al., 2004; Ogura et al., 2015). The diversity and distribution of these phage-derived elements highlight their influence on horizontal gene flow, virulence potential, and the overall evolutionary trajectory of *E. coli* O157:H7.

Table 2
Chromosomal sequences examined using PHASTEST revealed phage-like components in strains of *E. coli* O157:H7.

Strain name	Prophages	Intact	Incomplete	Questionable	GC content %
Sakai substr. RIMD	22	13	5	4	49.86
1044	26	11	10	5	50,43
1125	27	12	11	4	50,32
EC508	30	20	7	3	50,65
EC536	10	2	3	5	50,24
EC869	29	19	6	4	50,54
EC1212	26	12	8	6	50,07
EC4009	7	2	3	2	50,22
EC4042	18	14	7	2	50,46
EC4045	23	19	1	3	50,47
EC4076	24	17	4	3	50,56
EC4084	7	2	1	4	50,22
EC4113	27	17	5	5	50,6
EC4115	17	15	1	1	50,41
EC4127	8	2	2	4	50,23
EC4191	9	2	3	4	50,22
EC4192	8	2	3	3	50,24
EC4196	26	14	6	6	50,55
EC4205	5	0	1	4	50,21
EC4206	20	18	1	1	50,47
EC4401	26	16	6	6	50,46
EC4486	25	14	6	5	50,42
EC4501	30	18	5	7	50,58
EDL933	15	13	1	1	50,43
F8092B	19	14	3	2	50,51
FRIK966	16	5	6	5	50,24
FRIK2000	16	6	6	4	50,18
G5101	11	3	3	5	50,59
G5303	10	5	3	2	50,13
H093800014	14	6	4	4	50,3
LSU-61	9	3	4	2	50,61
SS17	16	14	1	1	50,38
SS52	16	14	1	1	50,47
T1543_06	9	3	2	4	50,21
TW14313	10	2	5	3	50,45
TW14359	17	15	1	1	50,46
TW14588	17	16	1	0	50,45

3.5. Antimicrobial resistance genes

Screening for antimicrobial resistance (AMR) determinants identified six resistance genes, *catA1*, *aph(6)-Id*, *aph(3″)-Ib*, *floR*, *sul2*, and *tetA*, present in a subset of *E. coli* O157:H7 genomes included in this analysis (Fig. 2). Among these isolates, EC4192 carried the highest number of AMR genes, harboring five distinct resistance determinants, whereas EC4401 contained only *catA1*. No AMR genes were detected in the Sakai reference genome.

The distribution of AMR genes was limited and highly variable across isolates, indicating that antimicrobial resistance is not uniformly conserved within *E. coli* O157:H7. Instead, AMR determinants appeared in isolated cases and were primarily associated with strains showing greater divergence in mobile genetic element regions. This pattern is consistent with prior observations that resistance genes within O157:H7

Table 3
Chromosomal sequences examined using PHASTEST revealed phage-like components in strains of *E. coli* O157:H7.

Strain name	Prophages	Intact	Incomplete	Questionable	GC content %
06–3745	13	3	7	3	50,17
06–4039	11	2	6	3	50,13
07–3091	11	3	5	3	50,09
07–3391	12	3	4	5	50,14
08–3037	12	2	6	4	50,17
08–3527	13	3	6	4	50,16
08–4169	12	4	5	3	50,13
08–4529	13	4	5	4	50,13
08BKT61141	9	2	4	3	50,2
08BKT048303	9	2	4	3	50,21
2009C-4258	13	4	5	4	50,05
2009EL1449	13	5	6	2	50,22
2009EL1705	15	4	6	5	50,3
2009EL1913	13	5	3	5	50,35
2009EL2109	16	5	7	4	50,4
2010C-4979C1	6	2	3	1	50,1
2011EL-1107	16	4	6	6	50,36
2011EL-2090	12	3	5	4	50,17
2011EL-2091	12	4	5	3	50,19
2011EL-2092	11	4	4	3	50,21
2011EL-2093	12	4	5	3	50,2
2011EL-2094	12	4	5	3	50,21
2011EL-2096	11	4	4	3	50,22
2011EL-2097	12	3	5	4	50,18
2011EL-2098	11	4	4	3	50,21
2011EL-2099	13	3	6	4	50,2
2011EL-2101	11	4	4	3	50,19
2011EL-2103	12	3	4	5	50,14
2011EL-2104	11	3	4	4	50,12
2011EL-2105	13	3	6	4	50,12
2011EL-2106	13	3	5	5	50,11
2011EL-2107	12	3	4	5	50,11
2011EL-2108	13	3	6	4	50,11
K2188	16	5	7	4	50,37
K4396	18	5	10	3	50,29
K5453	14	3	7	4	50,26
K5806	13	6	4	3	50,28
K6590	16	4	7	5	50,27
K7140	11	3	5	3	50,14

are often acquired through horizontal gene transfer events mediated by plasmids or prophage-associated regions, rather than being core genomic features of this lineage (Doyle et al., 1997).

These findings support the notion that AMR gene presence in *E. coli* O157:H7 is sporadic and largely linked to the acquisition or loss of exogenous mobile elements, reflecting the dynamic and variable nature of accessory genome content among strains.

3.6. Heatmap clustering and overall genomic trends

Double-hierarchical clustering of strains versus accessory genetic characteristics (including plasmid replicons and virulence genes) using double-hierarchical clustering revealed unique patterns of co-occurrence and grouping of isolates (Fig. 2; Table 1). Clusters characterized by preserved core-virulence loci (such as LEE and core T3SS effectors, particularly *esp/tir/fimH* family genes, *terC*, *chuA*, and *nlpI*) constitute a prominent central group of strains demonstrating high overall resemblance in essential pathogenic features. Marginal clusters are primarily identified by accessory characteristics: strains that are enriched for multiple Inc. family replicons such as *IncFIB*, *IncFII*, I*ncFIA*, *pEC4115* tend to cluster together with isolates that possess plasmid-related virulence genes, such as *ehxA* and *katP*; smaller groups are defined by the presence of uncommon replicons (*IncB/O/K/Z*, *Col156*, *IncN*, *Col(BS152)*, *IncFII[pSE11]*) in conjunction with specific accessory virulence markers such as *cba*, *efa1*, or distinct non-LEE effectors.

The heatmap clearly illustrates a connection between plasmid possession and varying virulence content: isolates with broader plasmid collections generally exhibit higher numbers of accessory virulence genes, while genomes that lack these plasmid replicons display fewer accessory virulence loci. The presence of antimicrobial resistance genes is infrequent in this analysis, appearing as individual occurrences on the heatmap. In the supplementary data, only two genomes (EC4192 and EC4401) contain AMR determinants (EC4192: *aph(6)-Id*, *aph(3″)-Ib*, *floR*, *sul2*, *tetA*; EC4401: *catA1*), and these AMR markers do not cluster together but are found in isolation within those strains in the dendrogram.

Overall, the heatmap underscores the idea that plasmid replicons are significant indicators of accessory virulence gene abundance and that the co-occurrence of plasmids and virulence is a key element of genomic differentiation within this dataset. Using BRIG/ANI, PHASTEST, PlasmidFinder, VirulenceFinder, and targeted AMR screening, the dataset reveals two distinct and complementary trends. A highly conserved core genome. Most *E. coli* O157:H7 genomes closely align with the Sakai reference, demonstrating high genome coverage and elevated ANI for core regions. The majority of isolates surpass 98 % coverage/identity, indicating a stable and conserved backbone throughout the collection. Numerous strains, such as EC4115, EC4042, EC4045, and EC4501, exhibit nearly complete alignment with the Sakai chromosome.

4. Discussion

Gram-negative species are present worldwide. *E. coli*, particularly the strain O157:H7, has gained recognition as a significant foodborne pathogen linked to human illnesses due to its genome's continual evolution through mutation events and horizontal gene transfer (Naidoo and Zishiri, 2023; Riley et al., 1983; Zhang et al., 2007; Manning et al., 2008; Ogura et al., 2009; Eppinger et al., 2011), allowing for the divergence and adaptation of strains to infect carrier hosts, resulting in human diseases or enabling survival in external conditions (Carter et al., 2012; Sharma et al., 2017; Carter et al., 2014). Thus, understanding the genomic diversity and adaptability of the O157:H7 strains is essential for predicting the disease's severity, comprehending bacterial pathogenesis, identifying specific biomarkers, tracking origins, evaluating epidemiology, and developing vaccines. In this research, comparative genomics was used to analyze the chromosomal sequence of the *E. coli* strain O157:H7 (Sakai), intending to uncover its genetic and functional attributes in comparison to other extensively studied O157 strains. Available online bioinformatics tools were used to collect data on chromosomal homology, the existence of plasmids, virulence genes, resistance genes, and prophages in the O157:H7 reference strain (Sakai), as well as other pertinent strains (Carter et al., 2012; Sharma et al., 2017; Carter et al., 2014).

Circular maps were generated using BRIG to compare the *E. coli* O157:H7 Sakai reference genome with the other strains included in this study and to assess overall chromosomal similarity. Several strains, such as EC4196, TW14588, 1044, and H093800014, demonstrated particularly high similarity to the reference, exhibiting minimal missing regions and few gaps, which indicates limited genomic variation relative to Sakai (Table 1). In contrast, a substantial proportion of the genomes showed more extensive gaps or regions of reduced homology, reflecting greater divergence from the reference strain. Variations within these chromosomal regions are consistent with differences driven by mobile genetic elements, including prophages and genomic islands, which are known to influence virulence, antimicrobial resistance, and other phenotypic characteristics (Naidoo & Zishiri, 2023; Ogura et al., 2009). Although all *E. coli* share a core genome of approximately 4.1 Mb, pathogenic lineages such as O157:H7 frequently incorporate additional mobile DNA phages, genomic islands, and transposons that expand genome size and contribute to strain-specific diversity (Ogura et al., 2009; Sharma et al., 2019; Ohnishi et al., 2001; Perna et al., 2001). In pathogenic bacteria, antimicrobial resistance genes also play an

important role in survival against therapeutic agents, with mechanisms such as efflux systems mediating reduced susceptibility to antimicrobials (Miryala and Ramaiah, 2019; Naidoo and Zishiri, 2023).

Table 1 and Figs. 1–2 show that a subset of isolates exhibits notably high sequence conservation relative to the Sakai reference strain, with Average Nucleotide Identity (ANI) values frequently surpassing 98–99 % and genome coverage above 95 %. These strains retained essential virulence determinants, including the Locus of Enterocyte Effacement (LEE) pathogenicity island (1.2–1.4 Mbp), which was present across all examined genomes, reflecting its critical role in adherence and colonization (Doe et al., 2020). Shiga toxin (*Stx*) prophage regions, which contribute substantially to pathogenic potential, also remained preserved among most isolates, consistent with the stability of these phage-encoded loci reported in earlier studies (Lee et al., 2019). Plasmid-associated virulence genes such as *ehxA*, *espP*, and *katP* were found variably distributed but were most consistently detected in strains that also carried *IncF*-family plasmid replicons.

BRIG analysis revealed recurrent regions of variability between approximately 1–2 Mb and 4–5.5 Mb (Table 1). These segments correspond to known genomic hotspots enriched in prophages and genomic islands and are characterized by frequent insertions, deletions, and recombination (Nguyen et al., 2023). Several strains displayed unique GC skew patterns within these variable loci, suggesting recent horizontal gene transfer or prophage excision events. Prophage profiling using PHASTEST further demonstrated considerable diversity, with prophage counts ranging from 6 to 30 regions per genome (Tables 2–3). Strains with higher prophage content generally showed greater variation in accessory regions, while strains with lower prophage numbers displayed more conserved genomic profiles. These findings reinforce the central role of mobile genetic elements in shaping the genomic diversity of *E. coli* O157:H7.

In contrast, several isolates such as 2011EL-2101, 2011EL-2099, 06–3745, and 09BKT048303_4541 exhibited lower genomic coverage (approximately 85–93 %) and pronounced divergence in regions associated with prophage integration and mobile elements. These genomes contained larger numbers of gaps or reduced identity zones within accessory regions, reflecting strain-specific patterns of gene loss, acquisition, or modification of mobile elements. Despite these differences, approximately 80–85 % of the genome remained conserved across all isolates. This conserved core included essential housekeeping genes, chromosomal maintenance loci, and major virulence determinants such as *Stx*, *eae*, and *tir*.

Across the dataset, GC content and GC-skew profiles were stable, indicating that chromosomal architecture is well maintained within this lineage. The divergence observed among strains therefore appears to arise primarily from variability in prophages, genomic islands, and plasmid-associated regions rather than from large-scale chromosomal rearrangement (Brown et al., 2012; Nguyen et al., 2023). Collectively, these observations highlight the interplay between a stable core genome and a dynamic accessory genome, driven by mobile elements that contribute substantially to the evolutionary trajectory of *E. coli* O157:H7. The present study identified a wide range of virulence genes across all *E. coli* O157:H7 strains analyzed (Fig. 4). Among the 53 virulence-associated genes detected (Fig. 2), eleven *aslA*, *ehuA*, *aea-g01-gamma*, *espA*, *espB*, *espY2 (000868321)*, *fimH*, *nleC*, *nlpI*, *terC*, and *tir* were consistently present in 100 % of the genomes examined. These universally conserved loci represent essential components of adhesion, secretion, and host-cell interaction mechanisms characteristic of the O157:H7 lineage.

In contrast, several accessory virulence genes displayed variable distribution across strains. Genes such as *astA*, *iss*, *nleB*, *espJ*, and *ompT* were detected only in a subset of isolates, while other loci, including *csgA*, *ehxA*, *etpD*, *fdeC*, *hylE*, *iha*, and *traT,* were found exclusively in different, non-overlapping subsets of strains. This pattern aligns with the broader accessory genome variability revealed through BRIG and PHASTEST analyses and reflects the influence of mobile genetic elements in shaping virulence gene content. The virulence genes identified span multiple functional categories, including adherence, iron acquisition, toxins, Shiga toxin components, and both LEE-encoded and non-LEE-encoded type III secretion system (T3SS) effectors. Among these, *tir* plays a central role as a T3SS effector that serves as the receptor for intimin, facilitating intimate adherence and the formation of attaching-and-effacing lesions typical of STEC infections. Additional factors, such as the *tccP* gene product, contribute to actin remodeling and are known to enhance the pathogenic profile of enterohemorrhagic *E. coli* (Noll et al., 2018; Zhang et al., 2021b; Liu et al., 2013; Liao et al., 2013). Overall, the virulence gene distribution observed across the dataset emphasizes the coexistence of a highly conserved pathogenic core and a variable accessory repertoire influenced by horizontal gene transfer and mobile genetic elements.

The *E. coli* O157:H7 strains examined in this study exhibited limited antimicrobial resistance, with AMR determinants detected in only two isolates (Fig. 2; Supplementary Information). Strain EC4401 carried a chromosomal and transposon-associated variant of the chloramphenicol acetyltransferase gene (*catA1*), while strain EC4192 harbored five AMR genes, including *aph(6)-Id* and *aph(3″)-Ib* (aminoglycoside resistance), floR (florfenicol resistance), *tetA* (tetracycline resistance), and *sul2* (sulfonamide resistance). These genes collectively encode mechanisms associated with drug modification, target protection, and efflux-mediated resistance, consistent with the known molecular strategies used by *E. coli* to withstand antimicrobial exposure. Efflux-mediated resistance was also reflected by the presence of *mdfA*, a predicted membrane transporter belonging to the major facilitator superfamily (MFS). The *mdfA* gene encodes a 410-amino acid protein capable of exporting a broad range of cationic and zwitterionic lipophilic compounds, including tetracycline, rifampin, ethidium bromide, and various aminoglycosides (Naidoo and Zishiri, 2023; Paramita et al., 2020). Previous studies have shown that *MdfA* contributes to decreased susceptibility to fluoroquinolones, erythromycin, chloramphenicol, and multiple other antimicrobial classes (Edgar and Bibi, 1997).

Among the 76 genomes evaluated, only one strain exhibited a set of AMR genes consistent with multidrug resistance, while the remaining isolates lacked detectable resistance determinants. This pattern indicates that antimicrobial resistance within the dataset is sporadic and strain-specific, rather than widely distributed across the *E. coli* O157:H7 genomes analyzed. As observed in prior studies, even low-frequency acquisition of AMR genes can contribute to broader concerns regarding resistance development (Naidoo and Zishiri, 2023; Rijavec et al., 2006). Resistance in bacteria can arise through the acquisition and exchange of genetic material, particularly via horizontal gene transfer mechanisms such as conjugation, transformation, and transduction. Antimicrobials are widely used in agriculture, human medicine, and veterinary practice, and the misuse or excessive application of these compounds contributes significantly to the emergence and spread of resistance (Munita and Arias, 2016; Meek et al., 2015; Alonso et al., 2017; Naidoo & Zishiri, 2013). Plasmids are especially important in this process, as they commonly carry genes associated with antimicrobial resistance and virulence and can mediate their transfer between bacterial cells (Yang et al., 2015; Villa et al., 2010).

Among the eleven plasmid replicons identified across the genomes in this study (Figs. 2–3), the I*ncF* family comprising *IncFIA*, *IncFIB*, and *IncFII* was the most prevalent, consistent with prior reports highlighting its central role in *E. coli* pathogenicity and adaptation (Villa et al., 2010; Johnson and Nolan, 2009). *IncF* plasmids possess replication and maintenance systems supported by toxin–antitoxin modules and often encode multiple replicon types simultaneously, most commonly combinations of *FII* with *FIA* and/or *FIB* (Lambrecht et al., 2018). Several plasmids detected in the dataset were present only in a limited number of isolates, including *pEC4115*, *IncI1-I(Alpha)*, *IncB/O/K/Z*, *Col156*, *IncI2(Delta)*, *IncN*, *Col(BS152)*, and *IncFII(pSE11)* (Fig. 3). Their restricted distribution indicates variability within the accessory genome and suggests that only a subset of strains has acquired these additional

replicons.

Strain EC4192 carried both *IncFII* and *IncFIB* (Fig. 3), a plasmid combination described by Lambrecht et al. (2018) as common among multidrug-resistant *E. coli*. Consistent with this, EC4192 exhibited the broadest antimicrobial resistance profile in the dataset, harboring five AMR genes associated with resistance to aminoglycosides, florfenicol, tetracycline, and sulfonamides. This finding underscores the well-documented association between *IncF* plasmids and the carriage of diverse antimicrobial resistance determinants. Previous studies have noted that IncF plasmids may harbor genes conferring resistance to aminoglycosides, β-lactams, chloramphenicol, quinolones, and tetracyclines, further highlighting their role in shaping resistance phenotypes in *E. coli* (Noll et al., 2018; Naidoo and Zishiri, 2023). Although the pO157 plasmid has been extensively investigated in *E. coli* O157:H7, many of the additional plasmids detected in this study remain less well characterized in the literature. Nevertheless, the patterns observed here align with previous findings that *IncF* plasmids constitute a major reservoir of both virulence and resistance genes within this serotype and contribute significantly to its genomic plasticity and adaptation.

Rifampin is commonly used both therapeutically and prophylactically, including for the prevention of infections caused by *Staphylococcus* and *Neisseria meningitidis*, and it exhibits activity against a wide range of bacterial pathogens such as *E. coli* and *Pseudomonas* (Weinstein and Zaman, 2019). Clinically, resistance to rifampin can arise under conditions of poor adherence, pharmacokinetic variability, or inappropriate antimicrobial use (Weinstein and Zaman, 2019; Arnold et al., 2017). Experimental studies further demonstrate that exposure to sub-inhibitory antimicrobial concentrations can select for resistant mutants, highlighting the influence of low-level drug exposure on the emergence of antimicrobial resistance (Kohanski et al., 2010). Rifampin exerts its antibacterial effect by targeting the *rpoB* gene product, the β-subunit of the DNA-dependent RNA polymerase (Campbell et al., 2021; Weinstein and Zaman, 2019). Resistance typically results from point mutations within *rpoB*, which reduce rifampin binding affinity. Three regions of *rpoB* have been identified as mutation "hotspots," where amino acid substitutions are repeatedly observed in rifampin-resistant strains across multiple species (Farhat et al., 2013). Even a single amino acid change in one of these hotspot regions can confer a substantial level of resistance.

Cross-resistance, a phenomenon in which resistance to one antimicrobial confers reduced susceptibility to another, often occurs among drugs that share structural or mechanistic similarities (Pang et al., 2013; Weinstein and Zaman, 2019). For example, Oz et al. (2014) reported cross-resistance among gyrase inhibitors in *E. coli* exposed to a single drug, suggesting that selective pressure on one antimicrobial target can influence susceptibility patterns within related drug classes. Additional studies have indicated that exposure to a degraded by-product of a drug may similarly select for resistance to the parent antimicrobial (Pincock, 2003).

The management of antimicrobial resistance is further complicated by treatment challenges such as adherence issues, transmission dynamics, and the circulation of substandard or counterfeit medications. These products differ from quality-assured formulations in dosage accuracy, bioavailability, or stability, and may result from manufacturing deficiencies or improper storage conditions. Antimicrobials are among the medications most frequently affected by such quality failures, contributing to therapeutic inefficacy and the potential amplification of resistance (Johnston and Holt, 2014; van Crevel et al., 2004; Weinstein and Zaman, 2019). Polymyxins are polypeptide antimicrobials first isolated in 1947 from *Paenibacillus polymyxa* subsp. *colistinus* (Phan et al., 2017). Polymyxin B and colistin (polymyxin E) show potent activity against a broad spectrum of Gram-negative bacteria, including members of the *Enterobacteriaceae*. Despite their effectiveness, their clinical use was historically limited because of significant neurotoxic and nephrotoxic side effects (Wadia and Tran, 2014; Phan et al., 2017). The global rise of multidrug-resistant (MDR) Gram-negative pathogens, particularly carbapenem-resistant *Enterobacteriaceae*, has renewed interest in polymyxins as last-resort therapeutic agents (Falagas and Kasiakou, 2005; Falagas et al., 2008; Poirel et al., 2017; Phan et al., 2017).

The bactericidal action of polymyxins begins with their interaction with the negatively charged lipopolysaccharides (LPS) of the outer membrane, causing increased membrane permeability, destabilization of the inner membrane, and eventual cell lysis (Phan et al., 2017; Velkov et al., 2010). The predominant mechanism of resistance involves modifications of LPS that reduce its overall negative charge, thereby decreasing polymyxin binding affinity (Nikaido, 2003). These modifications are primarily regulated by the *PmrAB* two-component system (TCS), a key determinant of polymyxin resistance across organisms such as *Acinetobacter baumannii*, *E. coli*, *Klebsiella pneumoniae*, *Pseudomonas aeruginosa*, and *Salmonella enterica* (Olaitan et al., 2014; Phan et al., 2017).

The *pmrCAB* operon encodes the phosphoethanolamine transferase *PmrC* (also known as *EptA*), along with the regulatory proteins *PmrA* and *PmrB*. Mutations in *pmrA* or *pmrB* can constitutively activate *PmrA*, leading to increased expression of *PmrC* and the products of the *arnB-CADTEF-pmrE* operon, which mediate the addition of phosphoethanolamine and 4-amino-4-deoxy-L-arabinose to lipid A (Olaitan et al., 2014; Phan et al., 2017; Poirel et al., 2017). Additional pathways contributing to polymyxin resistance include alterations in the PhoPQ TCS, which upregulates *PmrAB* via *PmrD*; binding of polymyxins by surface polysaccharides in *K. pneumoniae* and *P. aeruginosa*; modification of Kdo with phosphoethanolamine in *E. coli*; and efflux-mediated mechanisms described in *K. pneumoniae*.

A particularly concerning development is the discovery of the plasmid-mediated *mcr-1* gene, now reported globally in numerous bacterial isolates (Padilla et al., 2010; Phan et al., 2017). *mcr-1* encodes a phosphoethanolamine transferase similar to PmrC and was first identified on a transferable IncI2 plasmid. Mobilization of this gene is facilitated by the ISApl1 composite transposon, and *mcr-1* has since been detected on multiple plasmid types as well as chromosomally integrated forms (Liu et al., 2015).

Polymyxin B and colistin remain critical therapeutic options for infections caused by uropathogenic *E. coli* (UPEC), a major cause of urinary tract infections. Within UPEC, certain globally disseminated multidrug-resistant lineages pose particular clinical challenges.

The clustering patterns observed for both plasmid profiles and virulence gene distributions indicate that *E. coli* O157:H7 populations are shaped by the interplay between a highly conserved core genome and a more variable accessory genome. The consistent presence of core virulence genes such as *Stx*, *eae*, and hemolysin-associated loci across the majority of isolates reflects their essential contribution to pathogenicity and agrees with previous findings showing that these elements are retained despite ongoing evolutionary pressures (Karmali et al., 2010; Gilmour et al., 2015). In contrast, variability in accessory virulence genes among certain strains demonstrates the influence of mobile genetic elements, niche-specific selection, and genomic rearrangements on strain-level diversity.

The clustering patterns also highlight the important role of plasmid content in differentiating isolates. Plasmids frequently encode both virulence and resistance determinants and serve as major drivers of horizontal gene transfer, facilitating the dissemination of adaptive traits within *E. coli* populations (Carattoli, 2011; Johnson and Nolan, 2009). The strong contribution of plasmid-associated features to cluster separation suggests that differences in plasmid composition may have a more substantial impact on strain divergence than variations in virulence gene profiles alone. Collectively, these observations emphasize that while *E. coli* O157:H7 possesses a stable set of core pathogenicity genes, and accessory genome components, particularly plasmids, play a pivotal role in shaping genomic diversity and the evolutionary trajectories of individual strains.

These findings are consistent with recent studies demonstrating that

E. coli O157:H7 possesses a mosaic accessory genome shaped by the activity of mobile genetic elements, including prophages, plasmids, and genomic islands (Dorneles et al., 2018; Ogura et al., 2017). Variation in plasmid content among the strains analyzed reflects the influence of diverse ecological and selective pressures, such as antimicrobial exposure, competition within microbial communities, and phage–bacterium interactions. The continuum observed from strains exhibiting highly conserved accessory elements to those with extensive variability suggests that the evolution of *E. coli* O157:H7 is largely driven by the selective gain or loss of mobile genetic components.

Patterns of prophage retention, plasmid carriage, and accessory gene distribution further support the view that mobile elements provide a flexible reservoir of genetic diversity. These components contribute to differences in virulence potential, genomic plasticity, and ecological adaptability within *E. coli* O157:H7 populations. Overall, the results reinforce the concept that prophages and plasmids play a central role in shaping the evolutionary trajectory of enterohemorrhagic *E. coli*, serving as dynamic sources of genetic innovation that influence both pathogenesis and environmental persistence.

CRediT authorship contribution statement

Sydney Menzeko Gambushe: Writing – original draft, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Peter Ayodeji Idowu:** Writing – review & editing, Visualization, Validation, Supervision, Resources, Project administration. **Oliver Tendayi Zishiri:** Writing – review & editing, Visualization, Validation, Supervision, Resources, Project administration, Funding acquisition, Data curation.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.plasmid.2025.102771.

Data availability

NCBI Databases

References

Abongo, S.O., Momba, M.N.B., 2009. Prevalence and potential link between *Escherichia coli O157:H7* isolated from drinking water and stool samples of diarrhoeic children in the Amathole District, South Africa. Water SA 35, 757–764.

Acar, J.F., Courvalin, P., Chabbert, Y.A., 1970. Methicillin-resistant staphylococcemia: bacteriological failure of treatment with cephalosporins. Antimicrob. Agents Chemother. 10, 280–285.

Alikhan, N.F., Petty, N.K., Ben Zakour, N.L., Beatson, S.A., 2011. BLAST ring image generator (BRIG): simple prokaryote genome comparisons. BMC Genomics 12 (1), 1–10.

Alonso, C., Zarazaga, M., Ben Sallem, R., Jouini, A., Ben Slama, K., Torres, C., 2017. Antibiotic resistance in *Escherichia coli* in husbandry animals: the African perspective. Lett. Appl. Microbiol. 64 (5), 318–334.

Anjum, M.F., Zankari, E., Hasman, H., 2018. Molecular methods for detection of antimicrobial resistance. *Antimicrobial resistance in bacteria from livestock and companion animals*, pp. 33–50.

Arimizu, Y., Kirino, Y., Sato, M.P., Uno, K., Sato, T., Gotoh, Y., Ogura, Y., 2019. Large-scale genome analysis of bovine commensal Escherichia coli reveals that bovine-adapted E. coli lineages are serving as evolutionary sources of the emergence of human intestinal pathogenic strains. Genome Res. 29 (9), 1495–1505.

Arndt, D., Grant, J.R., Marcu, A., Sajed, T., Pon, A., Liang, Y., et al., 2016. PHASTER: a better, faster version of the PHAST phage search tool. Nucleic Acids Res. 44 (W1), W16–W21.

Arnold, A., Cooke, G.S., Kon, O.M., Dedicoat, M., Lipman, M., Loyse, A., Butcher, P.D., Ster, I.C., Harrisonm, T.S., 2017. Drug-resistant TB: UK multicentre study (DRUMS): treatment, management, and outcomes in London and West midlands 2008–2014. J. Inf. Secur. 74, 260–271. https://doi.org/10.1016/j.jinf. 2016.12.005.

Bader, J.C., Lakota, E.A., Andes, D.R., et al., 2018. Time for precision: a world without susceptibility breakpoints. Open Forum Infect. Dis. 5, ofy282.

Barth, Stefanie A., Menge, Christian, Eichhorn, Inga, Semmler, Torsten, Lothar, H. Wieler, Pickard, Derek, Belka, Ariane, Berens, Christian, Geue, Lutz, 2016. The accessory genome of Shiga toxin-producing Escherichia coli defines a persistent colonization type in cattle. Appl. Environ. Microbiol. 82 (17), 5455–5464.

Bielaszewska, M., Prager, R., Kock, R., Mellmann, A., Zhang, W., Tschäpe, H., Tarr, P.I., Karch, H., 2007. Shiga toxin gene loss and transfer in vitro and in vivo during enterohemorrhagic Escherichia coli O26 infection in humans. Appl. Environ. Microbiol. 73 (10), 3144–3150.

Bortolaia, V., Kaas, R.S., Ruppe, E., Roberts, M.C., Schwarz, S., Cattoir, V., Philippon, A., Allesoe, R.L., Rebelo, A.R., Florensa, A.F., Fagelhauer, L., Chakraborty, T., Neumann, B., Werner, G., Bender, J.K., Stingl, K., Nguyen, M., Coppens, J., Xavier, B. B., Malhotra-Kumar, S., Westh, H., Pinholt, M., Anjum, M.F., Duggett, N.A., Kempf, I., Nykäsenoja, S., Olkkola, S., Wieczorek, K., Amaro, A., Clemente, L., Mossong, J., Losch, S., Ragimbeau, C., Lund, O., Aarestrup, F.M., 2020. ResFinder 4.0 for predictions of phenotypes from genotypes. J. Antimicrob. Chemother. 75 (12), 3491–3500.

Brockhurst, M.A., Harrison, E., Hall, J.P., Richards, T., McNally, A., MacLean, C., 2019. The ecology and evolution of pangenomes. Curr. Biol. 29 (20), R1094–R1103.

Brown, T., et al., 2012. Genomic adaptations of enterohemorrhagic *E. coli* in environmental reservoirs. Front. Microbiol. 12, 3456.

Brüssow, H., Canchaya, C., Hardt, W.D., 2004. Phages and the evolution of bacterial pathogens: from genomic rearrangements to lysogenic conversion. Microbiol. Mol. Biol. Rev. 68 (3), 560–602. https://doi.org/10.1128/MMBR.00036-04.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., et al., 2009. BLAST+: architecture and applications. BMC Bioinform. 10, 1–9.

Campbell, E.A., Korzheva, N., Mustaev, A., Murakami, K., Nair, S., Goldfarb, A., Darst, S. A., 2021. Structural Mechanism for Rifampicin Inhibition of Bacterial RNA Polymerase.

Carattoli, A., 2011. Importance of plasmids in the propagation of antimicrobial resistance. Int. J. Med. Microbiol. 301 (8), 716–727.

Carter, M.Q., Parker, C.T., Louie, J.W., Huynh, S., Fagerquist, C.K., Mandrell, R.E., 2012. RcsB contributes to the distinct stress fitness among *Escherichia coli* O157:H7 curli variants of the 1993 hamburger-associated outbreak strains. Appl. Environ. Microbiol. 78 (21), 7706–7719.

Carattoli, A., Zankari, E., García-Fernández, A., Voldby Larsen, M., Lund, O., Villa, L., Møller Aarestrup, F., Hasman, H., 2014. In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. Antimicrob. Agents Chemother. 58 (7), 3895–3903. https://doi.org/10.1128/AAC.02412-14. Epub 2014 Apr 28. PMID: 24777092; PMCID: PMC4068535.

Carter, M.Q., 2017. Decoding the ecological function of accessory genome. Trends Microbiol. 25 (1), 6–8.

Carter, M.Q., Louie, J.W., Huynh, S., Parker, C.T., 2014. Natural rpoS mutations contribute to population heterogeneity in *Escherichia coli* O157:H7 strains linked to the 2006 US spinach-associated outbreak. Food Microbiol. 44, 108–118.

Centers for Disease Control and Prevention (CDC), 1982. Isolation of *E. coli* O157:H7 from sporadic cases of hemorrhagic colitis—United States. MMWR Morb. Mortal. Wkly. Rep. 31, 580–585.

Chapman, P.A., Wright, D.J., Higgins, R., 1997. Untreated milk as a source of *Escherichia coli O157* infection in farm visitors. Epidemiol. Infect. 119, 299–305.

Courvalin, P., 1994. Transfer of antibiotic resistance genes between gram-positive and gram-negative bacteria. Antimicrob. Agents Chemother. 38 (7), 1447–1451.

Croxen, M.A., Finlay, B.B., 2013. Recent advances in understanding enteric pathogenic *Escherichia coli*. Clin. Microbiol. Rev. 26, 822–880. https://doi.org/10.1128/CMR.00022-13.

Davies, T.J., Stoesser, N., Sheppard, A.E., et al., 2020. Reconciling the potentially irreconcilable? Genotypic and phenotypic amoxicillin–clavulanate resistance in *Escherichia coli*. Antimicrob. Agents Chemother. 64, e02026–19.

Dobrindt, U., Hochhut, B., Hentschel, U., Hacker, J., 2004. Genomic islands in pathogenic and environmental microorganisms. Nat. Rev. Microbiol. 2 (5), 414–424.

Doe, J., et al., 2020. LEE island conservation and pathogenesis in E. coli O157:H7. Infect. Immun. 88 (6), e00212–20.

Dorneles, E.M.S., et al., 2018. Genomic diversity of *Escherichia coli* O157:H7 isolates using whole genome sequencing. Front. Microbiol. 9, 951.

Doyle, M.P., Zhao, T., Meng, J., Zhao, S., 1997. *Escherichia coli* O157:H7. J. Food Prot. 60, 1105–1121. https://doi.org/10.4315/0362-028X-60.10.1105.

Edgar, R., Bibi, E., 1997. MdfA, an *Escherichia coli* multidrug resistance protein with an extraordinarily broad spectrum of drug recognition. J. Bacteriol. 179 (7), 2274–2280.

Eppinger, M., Mammel, M.K., LeClerc, J.E., Ravel, J., Cebula, T.A., 2011. Genome signatures of *Escherichia coli* O157: H7 isolates from the bovine host reservoir. Appl. Environ. Microbiol. 77 (9), 2916–2925.

Falagas, M.E., Kasiakou, S.K., 2005. Colistin: the revival of polymyxins for the management of multidrug-resistant gram-negative bacterial infections. Clin. Infect. Dis. 40, 1333–1341.

Escobar-Páramo, P., Grenet, K., Le Menac'h, A., Rode, L., Salgado, E., Amorin, C., Gouriou, S., Picard, B., Rahimy, M.C., Andremont, A., Denamur, E., 2004. Large-scale population structure of human commensal Escherichia coli isolates. Applied and environmental microbiology 70 (9), 5698–5700.

Falagas, M.E., Grammatikos, A.P., Michalopoulos, A., 2008. Potential of old-generation antibiotics to address the current need for new antibiotics. Expert Rev. Anti-Infect. Ther. 6, 593–600.

Farhat, M.R., Shapiro, B.J., Kieser, K.J., Sultana, R., Jacobson, K.R., Victor, T.C., Warren, R.M., Streicher, E.M., Calver, A., Sloutsky, A., Kaur, D., Posey, J.E., Plikaytis, B., Oggioni, M.R., Gardy, J.L., Johnston, J.C., Rodrigues, M., Tang, P.K., Kato-Maeda, M., Borowsky, M.L., Muddukrishna, B., Kreiswirth, B.N., Kurepina, N., Galagan, J., Gagneux, S., Birren, B., Rubin, E.J., Lander, E.S., Sabeti, P.C., Murray, M., 2013. Genomic analysis identifies targets of convergent positive selection in drug-resistant *Mycobacterium tuberculosis*. Nat. Genet. 45, 1183–1189. https://doi.org/10.1038/ng.2747.

Ferdous, M., Zhou, K., Mellmann, A., Morabito, S., Croughs, P.D., de Boer, Friedrich, A. W., 2015. Is Shiga toxin-negative Escherichia coli O157: H7 enteropathogenic or enterohemorrhagic Escherichia coli? Comprehensive molecular analysis using whole-genome sequencing. J. Clin. Microbiol. 53 (11), 3530–3538.
Fleckenstein, J.M., Hardwidge, P.R., Munson, G.P., Rasko, D.A., Sommerfelt, H., Steinsland, H., 2010. Molecular mechanisms of enterotoxigenic Escherichia coli infection. Microbes Infect. 12 (2), 89–98.
Fleckenstein, J.M., Kuhlmann, F.M., 2019. Enterotoxigenic Escherichia coli infections. Curr. Infect. Dis. Rep. 21 (3), 9.
Founou, L.L., Founou, R.C., Allam, M., Ismail, A., Essack, S.Y., 2018. Draft genome sequence of an extended-spectrum β-lactamase (CTX-M-15)-producing Escherichia coli ST10 isolated from a nasal sample of an abattoir worker in Cameroon. J. Glob. Antimicrob. Resist. 14, 68–69.
Gilmour, M.W., et al., 2015. Genetic and phenotypic characterization of O157:H7 isolates. J. Clin. Microbiol. 53 (2), 420–427.
Grad, Y.H., Godfrey, P., Cerquiera, G.C., Mariani-Kurkdjian, P., Gouali, M., Bingen, E., Shea, T.P., Haas, B.J., Griggs, A., Young, S., Zeng, Q., 2013. Comparative genomics of recent Shiga toxin-producing Escherichia coli O104: H4: short-term evolution of an emerging pathogen. MBio 4 (1), 10–1128.
Hendriksen, R.S., Bortolaia, V., Tate, H., et al., 2019. Using genomics to track global antimicrobial resistance. Front. Public Health 7, 242.
Herold, S., Karch, H., Schmidt, H., 2004. Shiga toxin-encoding bacteriophages–genomes in motion. Int. J. Med. Microbiol. 294 (2–3), 115–121.
Joensen, K.G., Scheutz, F., Lund, O., Hasman, H., Kaas, R.S., Nielsen, E.M., et al., 2014. Real-time whole-genome sequencing for routine typing, surveillance, and outbreak detection of verotoxigenic *Escherichia coli*. J. Clin. Microbiol. 52 (5), 1501–1510.
Johnson, T.J., Nolan, L.K., 2009. Pathogenomics of the virulence plasmids of *Escherichia coli*. Microbiol. Mol. Biol. Rev. 73 (4), 750–774.
Johnston, A., Holt, D.W., 2014. Substandard drugs: a potential crisis for public health. Br. J. Clin. Pharmacol. 78, 218–243. https://doi.org/10.1111/bcp.12298.
Juhas, M., 2019. Genomic Islands and the evolution of multidrug-resistant bacteria. *Horizontal Gene Transfer: Breaking Borders Between Living Kingdoms*, pp. 143–153.
Kaper, J.B., Nataro, J.P., Mobley, H.L., 2004. Pathogenic escherichia coli. Nat. Rev. Microbiol. 2 (2), 123–140.
Karch, H., Tarr, P.I., Bielaszewska, M., 2005. Enterohaemorrhagic *Escherichia coli* in human medicine. Int. J. Med. Microbiol. 295 (6–7), 405–418. https://doi.org/10.1016/j.ijmm.2005.06.009.
Karmali, M.A., et al., 2010. Pathogenesis of Shiga toxin–producing *Escherichia coli* infections. Clin. Microbiol. Rev. 23 (1), 39–88.
Kohanski, M.A., DePristo, M.A., Collins, J.J., 2010. Sublethal antibiotic treatment leads to multidrug resistance via radical-induced mutagenesis. Mol. Cell 37, 311–320. https://doi.org/10.1016/j.molcel, 01.0032010.
Kotloff, K.L., Nataro, J.P., Blackwelder, W.C., Nasrin, D., Farag, T.H., Panchalingam, S., Wu, Y., Sow, S.O., Sur, D., Breiman, R.F., 2013. Burden, and aetiology of diarrhoeal disease in infants and young children in developing countries (the global enteric Multicenter study, GEMS): a prospective, case-control study. Lancet 382, 209–222.
Lambrecht, E., Van Meervenne, E., Boon, N., Van de Wiele, T., Wattiau, P., Herman, L., et al., 2018. Characterization of cefotaxime-and ciprofloxacin-resistant commensal *Escherichia coli* originating from Belgian farm animals indicates high antibiotic resistance transfer rates. Microb. Drug Resist. 24 (6), 707–717.
Lee, S., et al., 2019. Shiga toxin phage diversity in clinical and environmental *E. coli* isolates. Microb. Genom. 5 (9), e000312.
Liao, X.-P., Liu, B.-T., Yang, Q.-E., Sun, J., Li, L., Fang, L.-X., et al., 2013. Comparison of plasmids coharboring 16S rRNA methylase and extended-spectrum β-lactamase genes among *Escherichia coli* isolates from pets and poultry. J. Food Prot. 76 (12), 2018–2023.
Liu, B.-T., Yang, Q.-E., Li, L., Sun, J., Liao, X.-P., Fang, L.-X., et al., 2013. Dissemination and characterization of plasmids carrying oqxAB-Bla CTX-M genes in *Escherichia coli* isolates from food-producing animals. PLoS One 8 (9), e73947.
Liu, Y.Y., Wang, Y., Walsh, T.R., et al., 2015. Emergence of plasmid-mediated colistin resistance mechanism MCR-1 in animals and human beings in China: a microbiological and molecular biological study. Lancet Infect. Dis. 16, 161–168.
Lockary, S., Besser, T.E., Hancock, D.D., McDermott, P.F., Wagner, B.A., Hill, W.E., 2007. Antimicrobial resistance among rectal *E. coli* isolates from healthy dairy cattle in the United States. Appl. Environ. Microbiol. 73, 1561–1565.
Lozano, R., Naghavi, M., Foreman, K., Lim, S., Shibuya, K., Aboyans, V., Abraham, J., Adair, T., Aggarwal, R., Ahn, S.Y., 2012. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the global burden of disease study 2010. Lancet 380, 2095–2128.
Lupolova, N., Lycett, S.J., Gally, D.L., 2019. A guide to machine learning for bacterial host attribution using genome sequence data. Microb. genom. 5 (12), e000317.
Malberg Tetzschner, A.M., Johnson, J.R., Johnston, B.D., Lund, O., Scheutz, F., 2020. In silico genotyping of *Escherichia coli* isolates for extraintestinal virulence genes by use of whole-genome sequencing data. J. Clin. Microbiol. 58 (10), e01269–20.
Manning, S.D., Motiwala, A.S., Springman, A.C., Qi, W., Lacher, D.W., Ouellette, L.M., et al., 2008. Variation in virulence among clades of *Escherichia coli* O157: H7 associated with disease outbreaks. Proc. Natl. Acad. Sci. 105 (12), 4868–4873.
Meek, R.W., Vyas, H., Piddock, L.J.V., 2015. Nonmedical uses of antibiotics: time to restrict their use? PLoS Biol. 13 (10), e1002266.
Mercer, R.G., Walker, B.D., Yang, X., McMullen, L.M., Gänzle, M.G., 2017. The locus of heat resistance (LHR) mediates heat resistance in Salmonella enterica, Escherichia coli and Enterobacter cloacae. Food Microbiol. 64, 96–103.
Miryala, S.K., Ramaiah, S., 2019. Exploring the multi-drug resistance in *Escherichia coli* O157:H7 by gene interaction network: a systems biology approach. Genomics 111 (4), 958–965.
Montero, D.A., Velasco, J., Del Canto, F., Puente, J.L., Padola, N.L., Rasko, D.A., Farfán, M., Salazar, J.C., Vidal, R., 2017. Locus of adhesion and autoaggregation (LAA), a pathogenicity island present in emerging Shiga toxin–producing Escherichia coli strains. Sci. Rep. 7 (1), 7011.
Mouton, J.W., Muller, A.E., Canton, R., et al., 2018. MIC-based dose adjustment: facts and fables. J. Antimicrob. Chemother. 73, 564–568.
Munita, J.M., Arias, C.A., 2016. Mechanisms of antibiotic resistance. Microbiol. Spectr. 4 (2), 4.2.15.
Naidoo, N., Zishiri, O.T., 2023. Comparative genomics analysis and characterization of Shiga toxin-producing Escherichia coli O157: H7 strains reveal virulence genes, resistance genes, prophages and plasmids. BMC genomics 24 (1), 791.
Nataro, J.P., Kaper, J.B., 1998. Escherichia coli diarrheagenic. Clin. Microbiol. Rev. 11 (1), 142–201.
Nguyen, T., et al., 2023. Prophage evolution and horizontal gene transfer in enterohemorrhagic *E. coli*. Front. Genet. 14, 11562.
Nikaido, H., 2003. Molecular basis of bacterial outer membrane permeability revisited. Microbiol. Mol. Biol. Rev. 67, 593–656.
Nisa, I., Zahra, R., Shabbir, M., Bhatti, M.F., 2013. Genomic diversity of *Escherichia coli O157:H7* isolates from Pakistan. J. Infect. Dev. Ctries. 7, 618–623.
Noll, L.W., Worley, J.N., Yang, X., Shridhar, P.B., Ludwig, J.B., Shi, X., et al., 2018. Comparative genomics reveals differences in mobile virulence genes of *Escherichia coli* O103 pathotypes of bovine fecal origin. PLoS One 13 (2), e0191362.
Ochman, H., Lawrence, J.G., Groisman, E.A., 2000. Lateral gene transfer and the nature of bacterial innovation. Nature 405 (6784), 299–304.
Ochoa, T.J., Contreras, C.A., 2011. Enteropathogenic *Escherichia coli* infection in children. Curr. Opin. Infect. Dis. 24, 478–483.
Ogura, Y., Ooka, T., Iguchi, A., Toh, H., Asadulghani, M., Oshima, K., et al., 2009. Comparative genomics reveals the mechanism of the parallel evolution of O157 and non-O157 enterohemorrhagic *Escherichia coli*. Proc. Natl. Acad. Sci. 106 (42), 17939–17944.
Ogura, Y., Mondal, S.I., Islam, M.R., Mako, T., Arisawa, K., Katsura, K., Ooka, T., Gotoh, Y., Murase, K., Ohnishi, M., Hayashi, T., et al., 2015. The Shiga toxin 2 production level in enterohemorrhagic *Escherichia coli* O157:H7 is correlated with the subtypes of toxin-encoding phage. Sci. Rep. 5, 16663. https://doi.org/10.1038/srep16663.
Ogura, Y., et al., 2017. Parallel evolution of highly pathogenic *Escherichia coli* O157:H7 in multiple lineages. Proc. Natl. Acad. Sci. 114 (7), 1801–1806.
Ohnishi, M., Kurokawa, K., Hayashi, T., 2001. Diversification of *Escherichia coli* genomes: are bacteriophages the major contributors? Trends Microbiol. 9 (10), 481–485.
Olaitan, A.O., Morand, S., Rolain, J.M., 2014. Mechanisms of polymyxin resistance: acquired and intrinsic resistance in bacteria. Front. Microbiol. 5, 643.
O'Neill, J., 2016. Tackling drug-resistant infections globally: final report and recommendations.
Oz, T., Guvenek, A., Yildiz, S., Karaboga, E., Tamer, Y.T., Mumcuyan, N., Ozan, V.B., Senturk, G.H., Cokol, M., Yeh, P., Toprak, E., 2014. Strength of selection pressure is an important parameter contributing to the complexity of antibiotic resistance evolution. Mol. Biol. Evol. 31, 2387–2401. https://doi.org/10.1093/molbev/msu191.
Padilla, E., Llobet, E., Domenech-Sanchez, A., et al., 2010. *Klebsiella pneumoniae* AcrAB efflux pump contributes to antimicrobial resistance and virulence. Antimicrob. Agents Chemother. 54, 177–183.
Pang, Y., Lu, J., Wang, Y., Song, Y., Wang, S., Zhao, Y., 2013. Study of the rifampin monoresistance mechanism in *Mycobacterium tuberculosis*. Antimicrob. Agents Chemother. 57, 893–900. https://doi.org/10.1128/AAC.01024-12.
Paramita, R.I., Nelwan, E.J., Fadilah, F., Renesteen, E., Puspandari, N., Erlina, L., 2020. Genome-based characterization of *Escherichia coli* causing bloodstream infection through next-generation sequencing. PLoS One 15 (12), e0244358.
Paton, J.C., Paton, A.W., 1998. Pathogenesis and diagnosis of Shiga toxin-producing Escherichia coli infections. Clin. Microbiol. Rev. 11 (3), 450–479.
Perna, N.T., Plunkett, G., Burland, V., Mau, B., Glasner, J.D., Rose, D.J., et al., 2001. Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7. Nature 409 (6819), 529–533.
Phan, M., Thi, N., Nhu, K., Achard, A.E.S., Forde, B.M., Hong, K.W., Chong, T.M., Yin, W., Chan, K., West, N.P., Walker, M.J., Paterson, D.L., Beatson, S.A., Schembri, M.A.J., 2017. Modifications in the pmrB gene are the primary mechanism for the development of chromosomally encoded resistance to polymyxins in uropathogenic *Escherichia coli*. J. Antimicrob. Chemother. 72, 2729–2736. https://doi.org/10.1093/jac/dkx204.
Pincock, S., 2003. WHO tries to tackle the problem of counterfeit medicines in Asia. BMJ 327, 1126. https://doi.org/10.1136/bmj.327.7424.1126-a.
Poirel, L., Jayol, A., Nordmann, P., 2017. Polymyxins: antibacterial activity, susceptibility testing, and resistance mechanisms encoded by plasmids or chromosomes. Clin. Microbiol. Rev. 30, 557–596.
Qadri, F., Svennerholm, A.M., Faruque, A.S.G., Sack, R.B., 2005. Enterotoxigenic *Escherichia coli* in developing countries: epidemiology, microbiology, clinical features, treatment, and prevention. Clin. Microbiol. Rev. 18, 465–483.
Rahimi, E., Kazemeini, H.R., Salajegheh, M., 2012. Detection of *Escherichia coli O157:H7* in some traditional cheeses in Iran. Food Control 24, 227–230.
Rijavec, M., Erjavec, M.S., Avguštin, J.A., Reissbrodt, R., Fruth, A., Križan-Hergouth, V., et al., 2006. High prevalence of multidrug resistance and random distribution of mobile genetic elements among uropathogenic *Escherichia coli* (UPEC) of the four major phylogenetic groups. Curr. Microbiol. 53 (2), 158–162.
Riley, L.W., Remis, R.S., Helgerson, S.D., McGee, H.B., Wells, J.G., Davis, B.R., et al., 1983. Hemorrhagic colitis is associated with a rare *Escherichia coli* serotype. N. Engl. J. Med. 308 (12), 681–685.

Sharma, V., Bayles, D., Alt, D., Looft, T., Brunelle, B., Stasko, J., 2017. Disruption of rcsB by a duplicated sequence in a curli-producing *Escherichia coli* O157: H7 results in differential gene expression in relation to biofilm formation, stress responses, and metabolism. BMC Microbiol. 17 (1), 1–23.

Robinson, C.M., Sinclair, J.F., Smith, M.J., O'Brien, A.D., 2006. Shiga toxin of enterohemorrhagic Escherichia coli type O157: H7 promotes intestinal colonization. Proc. Natl. Acad. Sci. 103 (25), 9667–9672.

Scallan, E., Hoekstra, R.M., Angulo, F.J., Tauxe, R.V., Widdowson, M.A., Roy, S.L., Griffin, P.M., 2011. Foodborne illness acquired in the United States—major pathogens. Emerg. Infect. Dis. 17 (1), 7.

Sharma, V.K., Akavaram, S., Schaut, R.G., Bayles, D.O., 2019. Comparative genomics reveals structural and functional features specific to the genome of a foodborne *Escherichia coli* O157:H7. BMC Genomics 20 (1), 1–18.

Tarr, P.I., Gordon, C.A., Chandler, W.L., 2005. Shiga-toxin-producing *Escherichia coli* and hemolytic uraemic syndrome. Lancet 365, 1073–1086.

van Crevel, R., Nelwan, R.H., Borst, F., Sahiratmadja, E., Cox, J., van der Meij, W., de Graaff, M., Alisjahbana, B., de Lange, W.C., Burger, D., 2004. Bioavailability of rifampicin in Indonesian subjects: a comparison of different local drug manufacturers. Int. J. Tuberc. Lung Dis. 8, 500–503.

Velkov, T., Thompson, P.E., Nation, R.L., et al., 2010. Structure-activity relationships of polymyxin antibiotics. J. Med. Chem. 53, 1898–1916.

Villa, L., Garcia-Fernandez, A., Fortini, D., Carattoli, A., 2010. Replicon sequence typing of IncF plasmids carrying virulence and resistance determinants. J. Antimicrob. Chemother. 65 (12), 2518–2529.

Wadia, S., Tran, B., 2014. Colistin-mediated neurotoxicity. BMJ Case Rep. 2014. https://doi.org/10.1136/bcr-2014-205332.

Weinstein, Z.B., Zaman, M.H., 2019. Evolution of rifampin resistance in *Escherichia coli* and *Mycobacterium smegmatis* due to substandard drugs. Antimicrob. Agents Chemother. 63, e01243–18. https://doi.org/10.1128/AAC.01243-18.

Werber, D., Fruth, A., Liesegang, A., Littmann, M., Buchholz, U., Prager, R., Voigt, W., Helms, M., Uphoff, H., Stark, K., 2007. A multistate outbreak of Shiga toxin-producing *Escherichia coli* O26:H11 infections in Germany was detected by molecular subtyping surveillance. J. Infect. Dis. 195, 358–365.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L.D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T.L., Miller, E., Bache, S.M., Müller, K., Ooms, J., Robinson, D., Seidel, D.P., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K., Yutani, H., 2019. Welcome to the tidyverse. J. Open Source Softw. 4 (43), 1686. https://doi.org/10.21105/joss.01686.

Yang, Q.-E., Sun, J., Li, L., Deng, H., Liu, B.-T., Fang, L.-X., et al., 2015. IncF plasmid diversity in multi-drug-resistant *Escherichia coli* strains from animals in China. Front. Microbiol. 6, 964.

Zankari, E., Allesøe, R., Joensen, K.G., et al., 2017. PointFinder: a novel web tool for WGS-based detection of antimicrobial resistance associated with chromosomal point mutations in bacterial pathogens. J. Antimicrob. Chemother. 72, 2764–2768.

Zhang, Y., Laing, C., Steele, M., Ziebell, K., Johnson, R., Benson, A.K., et al., 2007. Genome evolution in major *Escherichia coli* O157: H7 lineages. BMC Genomics 8 (1), 1–16.

Zhang, P., Essendoubi, S., Keenliside, J., Reuter, T., Stanford, K., King, R., et al., 2021a. Genomic analysis of Shiga toxin-producing *Escherichia coli* O157:H7 from cattle and pork-production related environments. npj Sci. Food 5 (1), 1–12.

Zhang, X., Pettengill, J.B., Lan, R., 2021b. Cluster-specific gene markers enhance *Shigella* and enteroinvasive *Escherichia coli* in-silico serotyping. Microb. Genom. 7 (11). https://doi.org/10.1099/mgen.0.000704.

Zhou, Y., Liang, Y., Lynch, K.H., Dennis, J.J., Wishart, D.S., 2011. PHAST: a fast phage search tool. Nucleic Acids Res. 39 (suppl_2), W347–W352.