ASIAN JOURNAL OF AGRICULTURE
Volume 9, Number 2, December 2025
Pages: 582-589

E-ISSN: 2580-4537
DOI: 10.13057/asianjagric/g090224

Sequence-based characterization of *Escherichia coli* in fresh produce and environmental samples in selected urban farms in Metro Manila, Philippines

PIERANGELI G. VITAL♥, MA. CHRISTINE JASMINE F. SABIO, DONNABEL C. SENA
Biological Research and Services Laboratory, Natural Sciences Research Institute, University of the Philippines Diliman. 1101 Quezon City, Philippines.
Tel.: +63-8981-8500 loc 3604, ♥email: pgvital@up.edu.ph

Manuscript received: 11 April 2025. Revision accepted: 30 September 2025.

Abstract. *Vital PG, Sabio MCJF, Sena DC. 2025. Sequence-based characterization of* Escherichia coli *in fresh produce and environmental samples in selected urban farms in Metro Manila, Philippines. Asian J Agric 9: 582-589.* The prevalence of *Escherichia coli* is well-documented in the Philippines, and this commonly food- and water-borne pathogen poses a public health risk. Serotyping is being done to determine the pathogenicity or clinical relevance of an isolate. With whole genome sequence (WGS) data becoming more available and the development of user-friendly online platforms for processing such data becoming more popular, an accurate strain identification can be made. The current study used web-based methods for multilocus sequence typing (MLST) and serotyping based on O and H antigens to characterize 30 *E. coli* isolates from a total of 117 samples collected from selected urban farms in the Philippines. Thirty-two out of 117 samples were confirmed to harbor thermotolerant *E. coli*. Sequence type 155 (ST 155), which is documented to be associated with zoonotic infections and antibiotic resistance, was observed among the isolates along with 25 other sequence types of *E. coli*. Several non-O157 Shiga toxin-producing *E. coli* (STEC) serotypes were detected, including O115:H11, O75:H10, O9a:H4, O148:H8, and O18ac:H21. Rare serotypes and incomplete serotypes (H-types only) were also detected. Our study highlights the utility of whole genome sequence-based strain characterization. The results of this study can be used for understanding contamination dynamics in the food chain and aiding agricultural food safety and outbreak investigations.

Keywords: *Escherichia coli*, MLST, sequence type 155, urban farms, whole genome sequencing

INTRODUCTION

Escherichia coli is a Gram-negative, facultatively anaerobic rod-shaped bacterium that belongs to the Enterobacteriaceae family. It is commonly found in the gastrointestinal tract of humans and animals, where most strains are harmless commensals (Bonten et al. 2021). However, certain *E. coli* strains have evolved pathogenic capabilities and can cause severe infections, including gastrointestinal, urinary, and systemic diseases, depending on the strain and virulence factors involved (Cabrera-Sosa and Ochoa 2020).

Escherichia coli plays an important ecological role and can be used as bioindicator of antimicrobial resistance. Animals used for food production carry this organism in the gastrointestinal tract making it a commensal; however, some strains remain the frequent cause of bacterial infections in humans and animals. Contamination with subsequent spread of antimicrobial resistance will affect the farm environment, from the source (including water, soil) up to the animals and humans.

In addition, the increasing emergence of multi-drug-resistant (MDR) *E. coli* strains, often due to the horizontal transfer of resistance genes via plasmids, poses a significant global public health threat (Kloos et al. 2021). Thus, effective strategies for monitoring the spread of these resistant pathogens are critical for safeguarding public health. Bacterial strain typing is a critical tool for identifying bacteria at the strain level and uncovering the genetic diversity underlying their behavior. It plays a central role in epidemiological surveillance, tracking outbreaks, understanding bacterial population structures, and informing clinical diagnosis and treatment strategies (Simar et al. 2021). One traditional approach for further classifying *E. coli* is serotyping, which targets surface antigens, specifically the O-polysaccharides (O-antigens) and H-flagellar proteins. However, more than 180 O-groups and 50 H-types have been documented, making phenotypic assays both labor-intensive and time-consuming, as they require induction of antigen expression under laboratory conditions (Jenkins 2015; Framatico et al. 2016).

Furthermore, a major limitation of these conventional methods is cross-reactivity, which can result in ambiguous or inaccurate serotype identification. For example, the O13 serogroup comprises structurally similar and cross-reactive serotypes (O13, O129, and O135) that vary only slightly due to the presence or absence of glucose residues and O-acetyl groups in their O-antigen structures. These antigens are also structurally related to those of *Shigella flexneri*, further complicating precise identification (Royer et al. 2022). The widely used agglutination-based serotyping methods are unable to differentiate between these closely related O-serotypes, highlighting the need for a more accurate and definitive typing tool (DebRoy et al. 2017).

In contrast to traditional phenotypic methods, genotypic approaches offer faster, more precise, and high-resolution

alternatives for bacterial subtyping. Earlier, pulsed-field gel electrophoresis (PFGE) became the gold standard for bacterial typing, allowing the generation of strain-specific DNA fingerprint patterns (Uelze et al. 2020). Later, one of the major advancements in genotypic methods was the advent of next-generation sequencing (NGS) platforms, which allowed a more cost-effective approach to producing whole-genome sequence (WGS) data. With WGS, a more comprehensive characterization of bacterial strains can be performed. WGS enables the in-silico prediction of O- and H-antigens, facilitating accurate serotype determination without the need for laborious antigen-induction assays (Grinevich et al. 2024). Later developments included platforms such as SerotypeFinder, which utilize mapping alignment to assign serovar identities from sequencing reads by matching them to known antigen gene databases (Uelze et al. 2020).

Additionally, whole-genome multilocus sequence typing (wgMLST), based on the sequence data of multiple internal fragments of housekeeping genes, allows classification of bacterial isolates at a fine resolution, defined as sequence types (STs) (Nakamura et al. 2021). MLST of *E. coli* isolates is done based on sequences from seven key housekeeping genes: *adk* (adenylate kinase), *fumC* (fumarate hydratase), *gyrB* (DNA gyrase), *icd* (isocitrate dehydrogenase), *mdh* (malate dehydrogenase), *purA* (adenylosuccinate synthetase), and *recA* (recombinase A). These genes are considered essential and stably expressed across various environmental conditions and cell states (Joshi et al. 2022). Their conserved yet polymorphic nature makes them ideal markers for strain differentiation (Decano and Downing 2019). The emergence of web-accessible genomic databases makes sequence typing and in-silico serotyping accessible even to laboratories with limited bioinformatics infrastructure, thus streamlining the subtyping process and enhancing our understanding of microbial diversity and virulence (Nouws et al. 2020; Uelze et al. 2020).

In this study, multilocus sequence typing (MLST) of *Escherichia coli* isolates and in-silico prediction of *E. coli* serotypes using WGS data were conducted. The bacterial isolates analyzed in this study were obtained from various samples including, vegetables, soil, water, and feces, collected from urban farms in Metro Manila, Philippines. These farms emerged during the COVID-19 pandemic as a response to growing concerns over food security and urban sustainability. This is the first comprehensive study done in Metro Manila, with emphasis on the farms that came about during the pandemic. Other studies did not include MLST or advanced molecular techniques. This will be essential in understanding contamination and antimicrobial resistance in the surrounding areas and in the entire country.

MATERIALS AND METHODS

Collection and maintenance of *Escherichia coli* strains

The sampling sites were urban farms located in Metro Manila which were established during the peak of COVID-19 pandemic, where community lockdowns and quarantines were frequently implemented, which often resulted to food scarcity, thereby leading to popularity of backyard farming (Vital et al. 2024). Three urban gardens were selected as sampling sites. For each site, vegetables were collected depending on the availability. Three vegetable samples were obtained from three spots in a plot, and these were combined to be considered as one sample only. Irrigation water and soil samples were also collected and processed. A total of 117 samples were obtained. Sampling was conducted during the period of October 2021 to January 2022. Thermotolerant *E. coli* isolation was performed using membrane filtration method and mTEC agar (Millipore, USA) (Vital et al. 2024). Incubation conditions used were 37°C for 2 h and then incubated at 44.5°C for 22 h. Presumptive *E. coli* (deep violet colonies) were streaked into Eosin Methylene Blue (EMB) (BD, Germany) agar plates and incubated overnight at 37°C. Typical *E. coli* colonies (green to black with or without green metallic sheen) were transferred to trypticase soy broth (TSB) (BD, Germany) and incubated at 37°C for 24 h. The resulting *E. coli* isolates in TSB were then subjected to DNA extraction. *Escherichia coli* ATCC 25922 served as positive control and NTC (No Template Control) as negative control. All experiments were done in triplicate.

DNA extraction and whole genome sequencing

Overnight cultures in TSB were subjected to DNA extraction using G-spin™ Total DNA Extraction Mini Kit (iNtRON Biotechnology, South Korea). Then, confirmation of *E. coli* was performed using polymerase chain reaction of *uidA* gene using primers ECN 1254 F (GCAAGGTGCACGGGAATATT) and ECN 1328 R (CAGGTGATCGGACGCGT) and GoTaq Green Mastermix (Promega, USA) in a reaction mix containing 1 µL of 1 mM forward primer 1 µL of 1 mM reverse primers, 1 µL of DNA template, and 4.5 µL of nuclease-free water, resulting in a total of 20 µL reaction mix. The PCR conditions used were as follows: initial denaturation at 98°C for 2 min, followed by 35 cycles of 95°C for 30 seconds, 63°C for 1 minute, and 72°C for 1 minute. The final extension was performed at 72°C for 1 minute. The positive control used for each run was *E. coli* ATCC 25922 (American Type Culture Collection). Visualized using gel electrophoresis, isolates that produced amplicons of about 75 base pairs were confirmed to be *E. coli.* DNA extracts of confirmed *E. coli* isolates were sent to Macrogen, Inc. (South Korea) for whole genome sequencing using Illumina HiSeq Platform.

Quality control and de novo assembly

Whole genome sequence data cleaning and quality control were performed using R programming and RStudio software. Data cleaning involved the removal of adapter sequences, N-bases, and low-quality paired-end reads. The de novo assembly and gene annotation were conducted using the Read Assembly and Annotation Pipeline Tool (RAPT) (National Center for Biotechnology Information).

Molecular characterization using MLST and construction of phylogenetic tree

MLST was performed on 30 samples using MLST v2.0.9 (https://cge.food.dtu.dk/services/MLST/) to determine the ST types (Wirth et al. 2006). On the MLST platform, the MLST configuration was set to "*Escherichia coli* #1" which uses the 5 housekeeping genes *adk*, *fumC*, *gyrB*, *icd*, *mdh*, *purA*, and *recA* as references. Minimum depth for an allele option was disabled since the input data used was pre-assembled. *E. coli* isolates were subjected to REALPHY v1.13 (The Swiss Institute of Bioinformatics) for a reference sequence alignment (Bertels et al. 2014). The neighbor-joining tree algorithm was used to construct the phylogenetic tree in the Molecular Evolutionary Genetics Analysis (MEGA) version 11 software (Tamura et al. 2021).

In-silico serotyping using SerotypeFinder

Serotype was identified using a web-based tool called SerotypeFinder 2.0 (htttps://cge.food.dtu.dk/services/SerotypeFinder/). Threshold percent identity was set to 85%. The minimum length or the number of nucleotides a sequence must overlap a serotype gene to count as a hit for that gene was set to 60% (Joensen et al. 2015).

RESULTS AND DISCUSSION

Thirty-two out of 117 samples were confirmed to harbor thermotolerant *E. coli*. Table 1 summarizes the number of positive samples per matrix. Two of the isolates were not processed for whole genome sequencing; hence, only 30 samples were subjected to MLST and in-silico Serotyping of the 30 samples analyzed, 25 samples were found to belong to a specific sequence type (ST) (Table 2).

The sequence types (STs) of the five *E. coli* isolates listed in Table 3 could not be determined. This means that the alleles for the housekeeping genes exhibited by these isolates did not match perfectly with the alleles registered in the database used for it to be considered as similar to a specific sequence type (Larsen et al. 2012). Meanwhile, although *E. coli* isolate V12UG3W had >99% coverage across all housekeeping genes, its sequence type remains unassigned. The closest match for this isolate was ST 1795.

Meanwhile, seven other isolates were assigned to ST155 (see Table 4).

Table 1. Number of *E. coli* isolates per sample matrix

Sample matrix	*E. coli* positive	*E. coli* negative	% Positive
Soil	2	16	11.11
Domestic animal feces	10	1	90.91
Fresh produce	15	56	21.13
Water	5	12	29.41

Table 2. Sequence types of *E. coli* isolates from various samples from selected urban farms in Metro Manila, Philippines

Sample type	Site 1	Site 2	Site 3
Water	ST69 ST1727	-	ST2628
Soil	-	ST10953	-
Fecal	ST2690 ST3580	ST196 ST211	-
Vegetables	ST2329 ST11231 ST101 ST2040; ST8410	ST1421 ST10 ST8262 ST29 ST2080	-

Table 3. *E. coli* isolates with unknown sequence type (ST) and percent coverage per locus of housekeeping genes

Sample code	% Similarity to housekeeping gene alleles registered in MLST database							Sample type	Source
	adk	*fumC*	*gyrB*	*icd*	*mdh*	*purA*	*recA*		
F2UG2D	100	100	100	100	100	94.77	100	Fecal	Site 2
S1UG2D	none	100	100	100	99.34	82.64	100	Soil	Site 2
V2UG2D	100	100	100	100	100	88.08	None	Vegetable	Site 2
V5UG2D	100	100	100	81.27	84.73	100	100	Vegetable	Site 2
V12UG3W	100	100	100	100	100	100	100	Vegetable	Site 3

Table 4. *E. coli* isolates with sequence type 155 and percent coverage per locus of housekeeping genes

Sample code	% Similarity to housekeeping gene alleles registered in MLST database							Sample type	Source
	adk	*fumC*	*gyrB*	*icd*	*mdh*	*purA*	*recA*		
F1UG1W	100	100	100	100	100	100	100	Fecal	Site 1
V6UG1W	100	100	100	100	100	100	100	Vegetable	Site 1
W1UG2W	100	100	100	100	100	100	100	Water	Site 2
W1UG2D	100	100	100	100	100	100	100	Water	Site 2
F1UG2W	100	100	100	100	100	100	100	Fecal	Site 2
F3UG2W	100	100	100	100	100	100	100	Fecal	Site 2
V10UG3W	100	100	100	100	100	100	100	Vegetable	Site 3

Table 5. Serotypes of *E. coli* isolates from various samples from selected urban farms in Metro Manila, Philippines

Sample type	Dry season			Wet season		
	Site 1	Site 2	Site 3	Site 1	Site 2	Site 3
Water	O17:H18	H34	-	H7	O109:H12 O153:H9	-
Soil		O156:H30	-		O156:H30	-
Fecal	O88:H16	O8:H51 H30 O22:H7	-	O28ac/O42:H21 O8:H12	O86:H51 O37:H10	-
Vegetables	O75:H10 O18ac:H21	O115:H11 O13:H7 O9a:H4	-	O88:H40 O159:H20 O22:H55 O159:H25	O5:H32 O114:H23 O115:H11 O96:H16	O148:H8 O37:H9

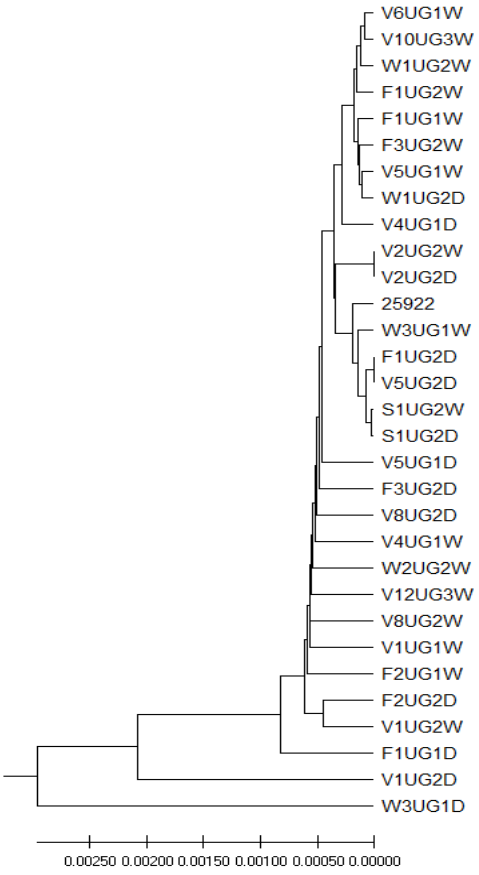

Figure 1. A neighbor-joining phylogenetic tree constructed using the MLST profiles of 14 *E. coli* isolates from selected urban farms in Metro Manila, Philippines

A phylogenetic tree was constructed using the neighbor-joining method in MEGA, with *E. coli* ATCC 25922 as the reference strain (Figure 1). Most isolates exhibited >99% sequence similarity, clustering closely at a 0.00100 distance, except for the *E. coli* isolates V1UG2D and W3UG1D (Figure 1). The results suggest limited genetic variation among *E. coli* isolates despite originating from different sample types or sampling locations. However, a larger sample size is needed to see more variation in the phylogenetic tree itself.

Finally, using SerotypeFinder, a total of 28 serotypes were identified from the input sequences (Table 5). The dataset includes both complete serotypes (O:H) and those where only H-types were determined (e.g., H30, H34, H7). The identified serotypes span a broad range of O and H types, indicating genetic diversity among the isolates. Notable serotypes include the non-O157 STEC-related serotypes.

Discussion

ST diversity and ST155 significance

The presence of distinct sequence types in the sample set, especially within the same sampling sites indicates genetic diversity. The genetic diversity demonstrated in our analysis is in alignment with a previous study, which characterized *E. coli* in environmental samples (Lasalde et al. 2005). Denaturing-gradient gel electrophoresis (DGGE) of *uidA* amplicons from *E. coli* isolated from pristine soils of a single sampling site showed genetically heterogeneous *E. coli* isolates. DGGE, like MLST, is a molecular typing method used for discriminating between bacterial strains. Genetic heterogeneity is widely true among microbial populations in nature and is particularly true in *E. coli*. This genetic diversity is attributable to variations in nutrient availability and other environmental drivers, creating several opportunities for mutation and horizontal gene transfer (Schreiber and Ackermann 2020; Yang et al. 2020).

Using MLST, the sequence type 155 (ST155) is repeatedly observed in this study in various samples as well as sampling sites. *E. coli* ST155 is well documented in the literature and is highly associated with zoonotic infections and antibiotic resistance (Castellanos et al. 2017; Gomi et al. 2017; Matamoros et al. 2017). More recently, a colistin-resistant *E. coli* isolate from a clinical sample in Germany was found to be of the ST155 (Neumann et al. 2020). Colistin is a last-line treatment for Gram-negative bacteria. More alarming is the fact that this specific *E. coli* isolate carries a novel variant of the *mcr* gene, which confers resistance to colistin. Thus, identifying the association of this sequence type with other antibiotic resistance genes is epidemiologically relevant. In our study, ST155 *E. coli* isolates were obtained from both soil, water, fecal, and vegetable samples, which potentially indicates potential predominance of this sequence locally. The presence of dominant sequence types suggests evolutionary success and adaptability to different environments and hosts. Different

environmental conditions both in nature and in the host, as well as the presence of antibiotics cause pressure to an organism to either gain or lose some of its genes to adapt to the environment. This ability to adapt causes the fluidity of the pangenome of an organism. Mechanisms like horizontal gene transfer, homologous recombination, and genome reduction develop fitness traits in organisms (Singh et al. 2024). Even in commensal strains, frequent recombination in the core genome leads to highly specialized strains over time which can emerge as a pandemic clone, which is evidence by dominant sequence types of organisms (McNally et al. 2013). However, the prevalence of this sequence type in our data set does not guarantee that it is the predominant type in the study area, as limitations in this study include the small number of samples and the narrow range of sequence types analyzed. However, the implications of having this specific sequence type must be further investigated.

Unknown ST's and database gaps

Five (5) of the analyzed samples were not assigned a specific sequence type (unknown ST). The occurrence of unknown STs stems from a non-matching allelic profile compared to that of the sequence types existing in the database for that organism. This demonstrates possible gaps in current database, sequencing errors, or even novel alleles (Jolley et al. 2010; Zhao et al. 2022; Thomas et al. 2024).

Table 4 suggests variable sequencing success across different isolates and sample types. Most likely, incomplete or missing data for one or more housekeeping genes is responsible for the inability to assign known sequence types to these isolates. Housekeeping genes may be missing in sequencing due to poor DNA quality (Larsen et al. 2012). Degraded or fragmented DNA reduces the efficiency of amplifying longer or less abundant genes required for MLST analysis. Improved sequencing or repeated testing may be necessary, particularly for samples with multiple low or missing genes.

Serotype diversity and relevance

In MLST, the variations in the sequences of the housekeeping genes are given distinct allele numbers, and the resultant allele patterns are used to discriminate between the strains. MLST has been demonstrated in various studies to be an effective approach for describing the genetic diversity of bacterial isolates.

Using SerotypeFinder, certain *Escherichia coli* serotypes were repeatedly detected from isolates obtained from the same sampling site. Serotype O156:H30 was identified in isolates from different seasons, while O115:H11 was detected in isolates collected during the same season. The recurrence of these serotypes suggests environmental persistence or potential reservoirs within the sampling area that support sustained colonization or re-introduction over time (Pilch et al. 2023). The seasonal recurrence of O156:H30 may reflect adaptability of this serotype to varying environmental conditions, possibly supported by specific survival mechanisms or host interactions (Van Elsas et al. 2011). On the other hand, the spatial and temporal consistency of O115:H11 implies a more stable niche presence, which could be indicative of limited microbial diversity or a consistent contamination source. Such persistence may also indicate that the serotype has become "naturalized" in specific environments. Yu et al. (2024) discussed the concept of different *E. coli* ecotypes and indicated that certain strains of *E. coli* can become permanent members of natural microbial communities in environments such as soil and water, leading to consistent detection over time. These observations underscore the utility of whole-genome sequencing (WGS-based) serotyping in identifying epidemiologically significant trends in serotype distribution, particularly when integrated with tools like SerotypeFinder. As reported by Joensen et al. (2015), SerotypeFinder achieved high concordance with conventional methods, successfully predicting 560 out of 569 O types and 504 out of 508 H types, validating its reliability in such ecological surveillance.

The presence of isolates carrying only H antigens (e.g., H30, H34, H7) was also detected. This may arise due to challenges in detecting or assembling O antigen genes, a known limitation of current WGS-based serotyping pipelines. As Joensen et al. (2015) highlighted, SerotypeFinder failed to detect O-processing genes in 32 isolates and flagellin genes in one isolate, attributing this to either poor sequence quality or the presence of unrecognized antigenic variants.

Furthermore, Uelze et al. (2020) pointed out that O antigen identification is particularly complex, as it relies on multiple genes, and sequence divergence or absence in databases can hinder proper annotation. This limitation suggests the possibility that the H-only results observed in this study could be due to incomplete genomic data or novel O antigen gene variants not represented in the current database.

Rare serotypes such as O8:H12 and O28ac/O42:H21 were also identified in this study. These findings are of interest due to their limited documentation in prior literature, highlighting the potential for WGS-based serotyping to uncover less common or emerging serotypes in environmental or clinical samples. The capacity to detect such rare types demonstrates the advantage of in silico tools over conventional serotyping due to limited antisera availability or cross-reactivity.

However, it is also crucial to recognize the limitations inherent to database completeness. As noted by Joensen et al. (2015), certain serotypes (e.g., O14, O57, O188) were not included in the SerotypeFinder database, leading to unassigned O types. Thus, the detection of rare or isolates with incomplete serotypes in this study may reflect either genuine rarity or gaps in the current serotyping databases, further supporting calls by Uelze et al. (2020) for continuous updates and standardization.

Aside from the duplicated, incomplete, and rare serotypes, a wide range of serotypes were also detected which are related to various *E. coli* strains that caused outbreaks as well as strains carrying multidrug resistance and pathogenicity genes as demonstrated by previous studies.

The shiga-toxin producing *E. coli* serotype O37:H10 was demonstrated to harbor the locus of enterocyte effacement (LEE), a pathogenicity island encoding a type 3 secretion system or T3SS which mediates the adhesion of *E. coli* into the microvilli (Sim et al. 2021). In the study of Wang et al. (2024), it was shown that *E. coli* serotype O115:H11, which was isolated from patients with bloody diarrhea and hemolytic uremic syndrome (HUS) in Sweden, contained the gene coding for EspP, a serine protease autotransporter which contributes to the degradation of human coagulation factor V, causing the bleeding of mucosal lining.

Serotypes O96:H16 and O17:H18 detected in this study were also found in uropathogenic *Escherichia coli* strains (UPEC) from previous studies (Hernández-Chiñas et al. 2021). UPEC is the predominant causative agent of recurrent urinary tract infections (rUTIs), particularly in pediatric populations. UPEC's dominance as a uropathogen is linked to its origin in the intestinal microbiota, with periurethral contamination serving as a key risk factor for infection. UPEC is not only isolated from clinical samples but is also found in the environment. Serotype O17:H18 was detected in *E. coli* isolate from a water sample in this study. Similarly, Boczek et al. (2007) were able to isolate clonal group A (CGA) *E. coli*, specifically serotype O17:H18 from various sewage effluents across the United States. CGA is a significant cause of drug resistant UPEC infections, especially those resistant to trimethoprim-sulfamethoxazole (TMP-SMZ).

Serotype H30 and its subclones are said to be the clonal basis of antimicrobial resistance, specifically fluoroquinolone resistance in *E. coli* sequence type 131 (ST131), an emerging globally disseminated multidrug resistant strain causing extraintestinal infections. A survey of 367 ST131 isolates from the World Health Organization (WHO) Collaborating Centre for Reference and Research on *Escherichia* and *Klebsiella* (database for years 1957 to 2011) revealed that the progressive emergence of antimicrobial resistance in ST131 was not only a result of generalized temporal increase in resistance, but rather can be attributed mainly to the rise of H30 subclones in ST131 strains (Olesen et al. 2014).

Serotype O153:H9 found from water sample in this study was also detected in *E. coli* ST648 from cloacal swabs of wild magnificent frigatebirds (*Fregata magnificens*) from the Alcatrazes Archipelago, a remote and protected island ecosystem in Brazil (Ewbank et al. 2022). The said strain was reported to harbor numerous antimicrobial genes such as *bla*CTX-M-2, *bla*CMY-2, and *qnrB* as well as virulence and resistance determinants for disinfectants, heavy metals, and environmental stressors. Likewise, extended-spectrum beta-lactamase genes were also demonstrated in serotypes O5:H32 and O114:H23 according to previous studies (Coppola et al. 2022; Awosile et al. 2023; Wang et al. 2024).

The wide range of serotypes suggests a genetically diverse population, possibly derived from different ecological sources or exhibiting various pathogenic potentials. However, it is still important to consider that the presence of these antigen allele sequences does not always result to their phenotypic expressions leading to the specific phenotype. In some contexts, phenotypic validation is still warranted, especially in outbreak investigations.

Applications of WGS and MLST in surveillance

Epidemiological investigations using various subtyping methods were prompted by the increased incidence of antibiotic resistance, wherein molecular typing methods are being progressively used to characterize these AR pathogens (Gaiarsa et al. 2019). The unambiguity and transferability of the sequence data used in this method are its main advantages over other molecular typing methods (Almasian-Tehrani et al. 2021). Furthermore, with whole genome sequences of bacteria available, remarkable variations among strains of the same species can be identified, making it possible to track the source of contamination of bacteria isolated, such as in food manufacturing plants. Indeed, MLST has been increasingly used in characterization and source tracking of bacterial pathogens (Costache et al. 2020; Sulaiman et al. 2021; Ma et al. 2022). Specifically, MLST can be applied to track sources of foodborne pathogen contamination. In a study of Nakamura et al. (2021), sources of *E. coli* isolates from final food products of a manufacturing factory were successfully tracked using core genome MLST (cgMLST). In this study, it was shown that four of *E. coli* strains from intermediate and final food products belonged to the same genogroups as some of *E. coli* strains from environmental swab samples of the factory, therefore, the dynamics of contamination in the said location were identified. In another study, cgMLST-based phylogenetic analysis also tracked possible clonal transmission of three *Salmonella enterica* serotypes (Mbandaka, Indiana, and Kentucky) in broiler chicken production chain in Sichuan, China (Wang et al. 2021). Additionally, during the initial outbreak investigation of the 2012 and 2013 *E. coli* (STEC) outbreaks, two of the biggest outbreaks in Belgium for the last 10 years, PFGE failed to fully position two human isolates to the outbreak cluster, but this was resolved in a retrospective outbreak investigation using cgMLST, which could have enabled a clearer outbreak cluster definition of the isolates (Nouws et al. 2020). Furthermore, tools like SerotypeFinder have proven to be rapid, efficient, and largely consistent with conventional methods (Joensen et al. 2015), particularly when used in command-line mode for integration into other bioinformatics pipelines (Uelze et al. 2020). Despite these strengths, limitations such as occasional gene non-detection, dependency on database completeness, and difficulty distinguishing closely related serovars remain challenges to full adoption (Uelze et al. 2020). Therefore, continued benchmarking, database update, and methodological improvements are needed to address these gaps.

Limitations and challenges

There are several factors that a farm environment that influence contamination. This particularly includes pre-harvest to post-harvest factors such as irrigation water, soil, and the surrounding environment. In the sampling sites, it

was evident that farm and domestic animals were present that warranted the presence of *E. coli*. Further, agricultural irrigation water used by the communities involves using on-farm and surface waters that circulate *E. coli* together with other microorganisms. With the agricultural contamination, comes the spread of antimicrobial resistance. While the country has national standards on agricultural food safety, this should be fully enforced and presence of accurate methods, such as molecular must be performed.

In conclusion, our current study applied WGS-based characterization of local *E. coli* isolates and thereafter demonstrated *E. coli* isolates with varying and similar STs found in a sampling location (ST155) in an agricultural site, as well as detection of a wide range of serotypes and rare serotypes using an in-silico serotyping. WGS-based characterization bypasses the limitations of phenotypic assays, including cross-reactivity, reagent availability, and labor intensity. Thus, it provides a fast, reproducible, and scalable method for characterizing the serological profile of bacterial isolates. This shows the potential utility of WGS-based characterization of strains for understanding the dynamics of fecal contamination in the food chain and tracking sources of contamination, for food safety as well as epidemiological investigations of outbreaks. However, WGS-based surveillance of infectious disease outbreaks is still very limited in the Philippines. This stems from research and surveillance funds restrictions, limiting the capacity building for next generation sequencing technologies in the Philippines, as well as challenges posed by restricted capacity for sample coordination and storage between laboratories, leading to a small number of samples (Nouws et al. 2020). Such challenges must be addressed in order to explore the full potential of NGS and MLST in pathogen characterization and disease surveillance.

ACKNOWLEDGEMENTS

The authors acknowledge the Department of Agriculture - Bureau of Agricultural Research (DA-BAR) of the Philippine Government for supporting this research.

REFERENCES

Almasian-Tehrani N, Alebouyeh M, Armin S, Soleimani N, Azimi L, Shaker-Darabad R. 2021. Overview of typing techniques as molecular epidemiology tools for bacterial characterization. Cell Mol Biomed Rep 1 (2): 69-77. DOI: 10.55705/cmbr.2021.143413.1016.

Awosile B, Fritzler J, Levent G, Rahman MDK, Ajulo S, Daniel I, Tasnim Y, Sarkar S. 2023. Genomic characterization of fecal *Escherichia coli* isolates with reduced susceptibility to beta-lactam antimicrobials from wild hogs and coyotes. Pathogens 12 (7): 929. DOI: 10.3390/pathogens12070929.

Bertels F, Silander OK, Pachkov M, Rainey PB, Van Nimwegen E. 2014. Automated reconstruction of whole-genome phylogenies from short-sequence reads. Mol Biol Evol 31 (5): 1077-1088. DOI: 10.1093/molbev/msu088.

Boczek LA, Rice EW, Johnston B, Johnson JR. 2007. Occurrence of antibiotic resistant uropathogenic *Escherichia coli* clonal Group A in wastewater effluents. Appl Environ Microbiol 73 (13): 4180-4184. DOI: 10.1128/AEM.02225-06.

Bonten M, Johnson JR, van den Biggelaar AH, Georgalis L, Geurtsen J, de Palacios PI, Gravenstein S, Verstraeten T, Hermans P, Poolman JT. 2021. Epidemiology of *Escherichia coli* bacteremia: A systematic literature review. Clin Infect Dis 72 (7): 1211-1219. DOI: 10.1093/cid/ciaa210.

Cabrera-Sosa L, Ochoa TJ. 2020. *Escherichia coli* diarrhea, in Hunter's Tropical Medicine and Emerging Infectious Diseases. Elsevier: 481-485. DOI: 10.1016/B978-0-323-55512-8.00046-6.

Castellanos LR, Donado-Godoy P, León M, Clavijo V, Arevalo A, Bernal JF, Timmerman AJ, Mevius DJ, Wagenaar JA, Hordijk J. 2017. High heterogeneity of *Escherichia coli* sequence types harbouring ESBL/AmpC genes on IncI1 plasmids in the Colombian poultry chain. PLoS One 12 (1): e0170777. DOI: 10.1371/journal.pone.0170777.

Coppola N, Cordeiro NF, Trenchi G, Esposito F, Fuga B, Fuentes-Castillo D, Lincopan N, Iriarte A, Bado I, Vignoli R. 2022. Imported one-day-old chicks as trojan horses for multidrug-resistant priority pathogens harboring *mcr-9*, *rmtG*, and extended-spectrum β-lactamase genes. Appl Environ Microbiol 88 (2): e01675-21. DOI: 10.1128/AEM.01675-21.

Costache C, Kalmár Z, Colosi HA, Baciu AM, Opriş RV, Györke A, Colosi IA. 2020. First multilocus sequence typing (MLST) of *Giardia duodenalis* isolates from humans in Romania. Parasites Vectors 13: 1-11. DOI: 10.1186/s13071-020-04248-2.

DebRoy C, Fratamico PM, Yan X, Baranzoni G, Liu Y, Needleman DS, Tebbs R, O'Connell CD, Allred A, Swimley M, Mwangi M, Kapur V, Raygoza Garay JA, Roberts EL, Katani R. 2016. Comparison of O-antigen gene clusters of all O-serogroups of *Escherichia coli* and proposal for adopting a new nomenclature for O-Typing. PLOS ONE 11 (1): e0147434. DOI: 10.1371/journal.pone.0147434.

Decano AG, Downing T. 2019. An *Escherichia coli* ST131 pangenome atlas reveals population structure and evolution across 4,071 isolates. Sci Rep 9 (1): 17394. DOI: 10.1038/s41598-019-54004-5.

Ewbank AC, Fuentes-Castillo D, Sacristán C, Esposito F, Fuga B, Cardoso B, Godoy SN, Zamana RR, Gattamorta MA, Catão-Dias JL, Lincopan N. 2022. World Health Organization critical priority *Escherichia coli* clone ST648 in magnificent frigatebird (*Fregata magnificens*) of an uninhabited insular environment. Front Microbiol 13: 940600. DOI: 10.3389/fmicb.2022.940600.

Fratamico, PM, DebRoy C, Liu, Y, Needleman DS, Baranzoni GM, Feng P. 2016. Advances in molecular serotyping and subtyping of Escherichia coli. Front Microbiol 7: 644. DOI: 10.3389/fmicb.2016.00644.

Gaiarsa S, Bitar I, Comandatore F, Corbella M, Piazza A, Scaltriti E, Villa L, Postiglione U, Marone P, Nucleo E, Pongolini S, Migliavacca R, Sassera D. 2019. Can insertion sequences proliferation influence genomic plasticity? Comparative analysis of *Acinetobacter baumanii* sequence type 78, a persistent clone in Italian hospitals. Front Microbiol 10 (2080): 1-14. DOI: 10.3389/fmicb.2019.02080.

Gomi R, Matsuda T, Matsumura Y, Yamamoto M, Tanaka M, Ichiyama S, Yoneda M. 2017. Whole-genome analysis of antimicrobial-resistant and extraintestinal pathogenic *Escherichia coli* in river water. Appl Environ Microbiol 83 (5): e02703-16. DOI: 10.1128/AEM.02703-16.

Grinevich D, Harden L, Thakur S, Callahan B. 2024. Serovar-level identification of bacterial foodborne pathogens from full-length 16S rRNA gene sequencing. mSystems 9 (3): e00757-23. DOI: 10.1128/msystems.00757-23.

Hernández-Chiñas U, Chávez-Berrocal ME, Ahumada-Cota RE, Navarro-Ocaña A, Rocha-Ramírez LM, Pérez-del Mazo Y, Alvarado-Cabello M, Pérez-Soto G, León-Alamilla LA, Acevedo-Monroy SE, Esquiliano D, Raya-Rivera AM, Eslava CA. 2021. Prospective study in children with complicated urinary tract infection treated with autologous bacterial lysates. Microorganisms 9 (9): 1811. DOI: 10.3390/microorganisms9091811.

Jenkins C. 2015. Commentary: Whole-Genome sequencing data for serotyping *Escherichia coli*—it's time for a change! J Clin Microbiol 53 (8): 2402-2403. DOI: 10.1371/journal.pone.0307830.

Jolley KA, Maiden MC. 2010. BIGSdb: scalable analysis of bacterial genome variation at the population level. BMC Bioinformatics 11: 1-11. DOI: 10.1186/1471-2105-11-595.

Joensen KG, Tetzschner AMM, Iguchi A, Aarestrup FM, Scheutz F. 2015. Rapid and easy *in silico* serotyping of *Escherichia coli* isolates by use of whole-genome sequencing data. J Clin Microbiol 53 (8): 2410-2426. DOI: 10.1128/JCM.00008-15.

Joshi CJ, Ke W, Drangowska-Way A, O'Rourke EJ, Lewis NE. 2022. What are housekeeping genes? PLOS Compu Biol 18 (7): e1010295. DOI: 10.1371/journal.pcbi.1010295.

Kloos J, Gama JA, Hegstad J, Samuelsen Ø, Johnsen PJ. 2021. Piggybacking on niche adaptation improves the maintenance of multidrug-resistance plasmids. Mol Biol Evol 38 (8): 3188-3201. DOI: 10.1093/molbev/msab091.

Lasalde C, Rodriguez R, Toranzos GA, Smith HH. 2005. Heterogeneity of *uidA* gene in environmental *Escherichia coli* populations. J Water Health 3 (3): 297-304. DOI: 10.2166/wh.2005.041.

Larsen MV, Cosentino S, Rasmussen S, Friis C, Hasman H, Marvig RL, Jelsbak L, Sicheritz-Pontén T, Ussery DW, Aarestrup FM, Lund O. 2012. Multilocus sequence typing of total-genome-sequenced bacteria. J Clin Microbiol 50 (4):1355-61. DOI: 10.1128/jcm.06094-11.

Ma CY, Ihara M, Liu S, Sugie Y, Tanaka H. 2022. Tracking the source of antibiotic-resistant *Escherichia coli* in the aquatic environment in Shiga, Japan, through whole-genome sequencing. Environ Adv 8: 100185. DOI: 10.1016/j.envadv.2022.100185.

Matamoros S, Van Hattem JM, Arcilla MS, Willemse N, Melles DC, Penders J, Vinh TN, Thi Hoa N, Bootsma MC, Van Genderen PJ, Goorhuis A. 2017. Global phylogenetic analysis of *Escherichia coli* and plasmids carrying the mcr-1 gene indicates bacterial diversity but plasmid restriction. Sci Rep 7 (1): 15364. DOI: 10.1038/s41598-017-15539-7.

McNally A, Cheng L, Harris SR, Corander J. 2013. The evolutionary path to extraintestinal pathogenic, drug-resistant *Escherichia coli* is marked by drastic reduction in detectable recombination within the core genome. Genome Biol Evol 5 (4): 699-710. DOI: 10.1093/gbe/evt038.

Nakamura A, Takahashi H, Arai M, Tsuchiya T, Wada S, Fujimoto Y, Shimabara Y, Kuda T, Kimura B. 2021. Molecular subtyping for source tracking of *Escherichia coli* using core genome multilocus sequence typing at a food manufacturing plant. PLoS One 16 (12): e0261352. DOI: 10.1371/journal.pone.0261352.

Neumann B, Rackwitz W, Hunfeld KP, Fuchs S, Werner G, Pfeifer Y. 2020. Genome sequences of two clinical *Escherichia coli* isolates harboring the novel colistin-resistance gene variants mcr-1.26 and mcr-1.27. Gut Pathog 12: 1-7. DOI: 10.1186/s13099-020-00375-4.

Nouws S, Bogaerts B, Verhaegen B, Denayer S, Crombé F, De Rauw K, Piérard D, Marchal K, Vanneste K, Roosens NH, De Keersmaecker SC. 2020. The benefits of whole genome sequencing for foodborne outbreak investigation from the perspective of a national reference laboratory in a smaller country. Foods 9 (8): 1030. DOI: 10.3390/foods9081030.

Olesen B, Frimodt-Møller J, Leihof RF, Struve C, Johnston B, Hansen DS, Scheutz F, Krogfelt KA, Kuskowski MA, Clabots C, Johnson JR. 2014. Temporal trends in antimicrobial resistance and virulence-associated traits within the *Escherichia coli* Sequence Type 131 Clonal Group and its *H* 30 and *H* 30-Rx subclones, 1968 to 2012. Antimicrob Agents Chemother 58 (11): 6886-6895. DOI: 10.1128/AAC.03679-14.

Pilch H, Aulik N, Sockett D, Czuprynski CJ. 2023. Survival of bovine-associated serotypes of Salmonella enterica in bedding sand. JDS Commun 4 (2): 86-90. DOI: 10.3168/jdsc.2022-0305.

Royer G, Clermont O, Condamine B, et al. 2022. O-Antigen targeted vaccines against *Escherichia coli* may be useful in reducing morbidity, mortality, and antimicrobial resistance. Clin Infect Dis 74 (2): 364-366. DOI: 10.1093/cid/ciab458.

Schreiber F, Ackermann M. 2020. Environmental drivers of metabolic heterogeneity in clonal microbial populations. Curr Opin Biotechnol 62: 202-211. DOI: 10.1016/j.copbio.2019.11.018.

Sim EM, Kim R, Gall M, Arnott A, Howard P, Valcanis M, Howden BP, Sintchenko V. 2021. Added value of genomic surveillance of virulence factors in Shiga Toxin-producing *Escherichia coli* in New South Wales, Australia. Front Microbiol 12: 713724. DOI: 10.3389/fmicb.2021.713724.

Simar SR, Hanson BM, Arias CA. 2021. Techniques in bacterial strain typing: Past, present, and future. Curr Opin Infect Dis 34 (4): 339-345. DOI: 10.1097/QCO.0000000000000743.

Singh A, Rani PS, Bandsode V, Nyambero M, Qumar S, Ahmed N. 2024. Drivers of virulence and antimicrobial resistance in Gram-negative bacteria in different settings: A genomic perspective. Infect Genet Evol 124: 105666. DOI: 10.1016/j.meegid.2024.105666.

Sulaiman IM, Tang K, Segars K, Miranda N, Sulaiman N, Simpson S. 2021. Application of MALDI-TOF mass spectrometry, and DNA sequencing-based SLST and MLST analysis for the identification of *Cronobacter* spp. isolated from environmental surveillance samples. Arch Microbiol 203 (8): 4813-4820. DOI: 10.1007/s00203-021-02465-6.

Tamura K, Stecher G, Kumar S. 2021. MEGA11: molecular evolutionary genetics analysis version 11. Mol Biol Evol 38 (7): 3022-3027. DOI: 10.1093/molbev/msab120.

Thomas IV JC, Cartee JC, Hebrank K, St Cyr SB, Schlanger K, Raphael BH, Kersh EN, Joseph SJ. 2024. Emergence and evolution of mosaic penA-60 and penA-237 alleles in a *Neisseria gonorrhoeae* core genogroup that was historically susceptible to extended-spectrum cephalosporins. Front Microbiol 15: 1401303. DOI: 10.3389/fmicb.2024.1401303.

Uelze L, Grützke J, Borowiak M, Hammerl JA, Juraschek K, Deneke C, Tausch H, Malorny B. 2020. Typing methods based on whole genome sequencing data. One Health Outlook 2 (1): 3. DOI: 10.1186/s42522-020-0010-1.

van Elsas, JD, Semenov AV, Costa R, Trevors JT. 2011. Survival of *Escherichia coli* in the environment: Fundamental and public health aspects. ISME J 5 (2): 173-183. DOI: 10.1038/ismej.2010.80.

Vital PG, Rivera WL, Sena DC, Catapat CJC, Sabio CJF. 2024. Thermotolerant *Escherichia coli* contamination in vegetables from selected urban farms and wet markets in Metro Manila, Philippines at the height of COVID-19 pandemic. Asia-Pacific J Sci Technol 29 (03): 1-9.

Wang YT, Lei CW, Liu SY, Chen X, Gao YF, Zhang Y, Tang Y, Zhang A, Yang X, Wang HN. 2021. Tracking *Salmonella enterica* by whole genome sequencing of isolates recovered from broiler chickens in a poultry production system. Intl J Food Microbiol 350: 109246. DOI: 10.1016/j.ijfoodmicro.2021.109246.

Wang L, Hua Y, Bai X, Zhang J, Mernelius S, Chromek M, Frykman A, Hansson S, Matussek A. 2024. Prevalence and Characteristics of Plasmid-Encoded Serine Protease EspP in Clinical Shiga Toxin-Producing *Escherichia coli* Strains from Patients in Sweden. Microorganisms 12 (3): 589. DOI: 10.3390/microorganisms12030589.

Wirth T, Falush D, Lan R, Colles F, Mensa P, Wieler LH, Karch H, Reeves PR, Maiden MC, Ochman H, Achtman M. 2006. Sex and virulence in *Escherichia coli*: an evolutionary perspective. Mol Microbiol 60 (5): 1136-1151. DOI: 10.1111/j.1365-2958.2006.05172.x.

Yang DD, Alexander A, Kinnersley M, Cook E, Caudy A, Rosebrock A, Rosenzweig F. 2020. Fitness and productivity increase with ecotypic diversity among *Escherichia coli* strains that coevolved in a simple, constant environment. Appl Environ Microbiol 86 (8): e00051-20. DOI: 10.1128/AEM.00051-20.

Yu D, Stothard P, Neumann NF. 2024. Emergence of potentially disinfection-resistant, naturalized *Escherichia coli* populations across food- and water-associated engineered environments. Sci Rep 14 (1): 13478. DOI: 10.1038/s41598-024-64241-y.

Zhao X, Zhao H, Zhou Z, Miao Y, Li R, Yang B, Cao C, Xiao S, Wang X, Liu H, Wang J. 2022. Characterization of extended-spectrum β-lactamase-producing *Escherichia coli* isolates that cause diarrhea in sheep in Northwest China. Microbiol Spectrum 10 (4): e01595-22. DOI: 10.1128/spectrum.01595-22.