Enteroaggregative *Escherichia coli* is the predominant diarrheagenic *E. coli* pathotype among irrigation water and food sources in South Africa.

Matthew Aijuka[a]; Araceli E. Santiago[b]; Jorge A. Girón[b]; James P. Nataro[b] and Elna M. Buys[a]*

[a]Department of Consumer and Food Sciences, University of Pretoria, Republic of South Africa

[b]Child Health Center, Department of Pediatrics, School of Medicine, University of Virginia, Charlottesville, Virginia, USA

*Corresponding author at Department of Consumer and Food Sciences, University of Pretoria, Lynwood Rd, Pretoria, 0002, South Africa

Email addresses: aijumatt@yahoo.com (M. Aijuka), aes8j@hscmail.mcc.virginia.edu (A. E Santiago), jagiron@yahoo.com (J. A Girón), jpn2r@hscmail.mcc.virginia.edu (J. P Nataro), elna.buys@up.ac.za (E. M Buys)

ABSTRACT

Diarrheagenic *E. coli* (DEC) has been implicated in foodborne outbreaks worldwide and have been associated with childhood stunting in the absence of diarrhoea. Infection is extraordinarily common, but the routes of transmission have not been determined. Therefore, determining the most prevalent pathotypes in food and environmental sources may help provide better guidance to various stakeholders in ensuring food safety and public health and advancing understanding of the epidemiology of enteric disease. We characterized 205 *E. coli* strains previously isolated from producer distributor bulk milk (PDBM)(118), irrigation water (48), irrigated lettuce (29) and street vendor coleslaw (10) in South Africa. Enteropathogenic *E. coli* (EPEC), enterotoxigenic *E. coli* (ETEC), enteroaggregative *E. coli* (EAEC) and diffusely adherent *E. coli* (DAEC) were sought. We used PCR and partial gene sequencing for all 205 strains while 46 out of 205 that showed poor resolution were subsequently characterized using cell adherence (HeLa cells).

PCR and partial gene sequencing of *aatA* and/or *aaiC* genes confirmed EAEC (2%, 5 out of 205) as the only pathotype. Phylogenetic analysis of sequenced EAEC strains with *E. coli* strains in GenBank showing ≥80% nucleotide sequence similarity based on possession of *aaiC* and *aatA* generated distinct clusters of strains separated predominantly based on their source of isolation (food source or human stool) suggesting a potential role of virulence genes in source tracking. EAEC 24%, 11 out of 46 strains (PDBM=15%, irrigation water=7%, irrigated lettuce=2%) was similarly the predominant pathotype followed by strains showing invasiveness to HeLa cells, 4%, 2 out of 46 (PDBM=2%, irrigated lettuce=2%), among stains characterized using cell adherence.

Therefore, EAEC may be the leading cause of DEC associated food and water-borne enteric infection in South Africa. Additionally, solely using molecular based methods targeting virulence gene determinants may underestimate prevalence, especially among heterogeneous pathogens such as EAEC.

Key words: Enteroaggregative *E. coli*, Evolutionary relationship, Virulence genes, Cell adherence assay, Multiplex PCR, Foodborne pathogens

Abbreviations: DEC-diarrheagenic *E. coli*; EAEC-enteroaggregative *E. coli*; EIEC-enteroinvasive *E. coli*; ETEC-enterotoxigenic *E. coli*; EPEC-enteropathogenic *E. coli*; DAEC-diffusely adherent *E. coli*; AIEC-adherently invasive *E. coli*; DMEM-Dulbecco's Modified Eagle Medium; PBS-Phosphate Saline Buffer; MEGA- Molecular Evolutionary Genetics Analysis; NICD- National Institute of Communicable Diseases; TSB-Tryptone Soy Broth; UPEC-Uropathogenic *E. coli*

1. Introduction

Diarrheagenic *Escherichia coli* (DEC) has long been associated with foodborne illness and outbreaks worldwide, thereby posing a risk to global food safety and public health. *E. coli* pathotypes commonly associated with illness amongst varying age groups and geographical locations include enteropathogenic (EPEC), enterotoxigenic (ETEC), enterohaemorrhagic (EHEC), enteroaggregative (EAEC), diffusely adherent (DAEC) and enteroinvasive (EIEC) *E. coli* (Croxen et al., 2013; Kaper, 2005; Nataro and Kaper, 1998). However, recent studies have implicated less characterized strains in disease outbreaks such as EAEC producing shigatoxins (*E. coli* O104:H4 in 2011) that caused a large scale foodborne outbreak throughout Europe in 2011 (Rasko et al., 2011). Additionally, adherent and invasive *E. coli* (AIEC) has been linked to patients with Crohn's disease (Nash et al., 2010). Outbreaks such as that occurring in Germany in 2011 suggest risk of highly virulent pathotypes emerging, possibly driven by factors such as climate change (Carlton et al., 2016).

In addition to diarrhoea, EAEC has been associated with growth faltering in children in the absence of diarrhoea (Acosta et al., 2016) and molecular studies have suggested that up to 80% of children may harbour EAEC at a point in time (Platts-Mills et al., 2015).

Variation in epidemiology of diarrhoeal disease and associated pathotypes has been linked to geographical, temporal and climatic conditions complicating food safety and public health prevention and control initiatives especially in resource limited countries (Carlton et al., 2016). South Africa like many developing and Sub-Saharan African countries experiences a high incidence of diarrhoea (Tau et al., 2012) in infants and immune compromised adults, such as HIV-positive patients (Samie et al., 2007). Such illness would disproportionately affect the low income population, which makes up a large part of the urban and rural population (World Bank, 2014) due to inadequate waste disposal and sanitation facilities

(Baker et al., 2016). Within South Africa, The National Institute of Communicable Diseases (NICD) initiated a national surveillance program to monitor diarrheagenic pathotypes (Tau et al., 2012).

This initiative coupled with additional site specific studies from clinical specimens (Bisi-johnson et al., 2011; Samie et al., 2007) and food and environmental sources (Adefisoye and Okoh, 2016; Castro-Rosas et al., 2012; Farrokh et al., 2013; Gemmell and Schmidt, 2012; Newell et al., 2010) have helped shed light on the magnitude of the diarrheal disease burden associated with DEC as well as its prevalence within the environment.

Gaps remain in understanding the prevalence of DEC pathotype(s) in food and environmental sources over varying geographical, temporal and sample sources. Such information would help link observed infections/outbreaks with more specific food and environmental sources providing more informed guidance to food safety and public health interventions (Newell et al., 2010). The present study followed from two major previously concluded studies. The first study was initiated by The South African Department of Agriculture and Water Research Commission to help characterize the bacterial quality of South African irrigation water (Aijuka et al., 2015). A high prevalence of faecal indicators within milk sold by unregulated retailers alarmed the South African Dairy Development Agency which subsequently initiated the second study to characterize its microbiological quality (Ntuli et al., 2017, 2016).

We sought to determine the most prevalent DEC pathotypes associated with *E. coli* previously isolated from food sources and irrigation water collected in South Africa over varying geographical and temporal spans. Additionally, we sought to determine food sources associated with highest prevalence of DEC.

2. Materials and Methods

2.1. *Escherichia coli* isolate source

A total of 205 *E. coli* strains previously isolated from irrigation water, irrigated lettuce, producer distributor bulk milk (PDBM) and street vendor coleslaw were used in this study. This bacterial collection included 48 isolates from irrigation water (Aijuka et al., 2015; Aijuka, 2014), 29 from irrigated lettuce (Aijuka, 2014),

118 from milk (pasteurized and unpasteurized PDBM) sold by small holder sellers (producer-distributer) (Ntuli et al., 2016) and 10 from street-vendor coleslaw purchased in Pretoria City, South Africa. All isolates collected over these studies were stored at -80°C at the Department of Food Science, University of Pretoria, South Africa in Tryptone Soy Broth (TSB)(Biolab Diagnostics (Pty) Ltd, Midrand, South Africa) containing 30% glycerol (Sigma-Aldrich, St. Louis, MO, USA).

2.2. Isolate resuscitation and transportation

The isolates were regrown in TSB, incubated at 37°C overnight and transferred onto freshly prepared TSB agar slants in McCartney bottles. Slants were incubated at 37°C overnight and couriered to The Child Health Research Center, Department of Pediatrics, University of Virginia School of Medicine, Charlottesville, Virginia USA where all subsequent analyses were done. All isolates were resuscitated in Luria Broth Base, Miller's modification (LB; AmericanBio Inc, Natick, MA, USA) and subsequently grown on LB plates with respective overnight incubation at 37°C prior to any analysis.

2.3. Haemolysin production in *E. coli* isolates

All 205 isolates were grown on 5% blood agar (Hardy Diagnostics, Santa Maria, CA, USA), incubated at 37°C and checked for beta or alpha haemolysis as an indicator of the presence of hemolysins (Greene et al., 2015).

2.4. Molecular characterization of EPEC, ETEC, EAEC and DAEC

ETEC, EPEC and EAEC pathotypes were identified using a multiplex polymerase chain reaction (PCR) (Panchalingam et al., 2012). PCR targets (Table 1) included ETEC heat-labile (LT) and heat-stable (STh) enterotoxin genes, the EPEC intimin (*eae* gene) outer membrane protein adhesin and *bfpA*, the gene encoding the bundle forming pili (BFP); the EAEC plasmid-encoded gene *aatA*; and the EAEC chromosomally encoded *aaiC* locus. The gene targets are known virulence determinants of their respective pathogens (Nataro and Kaper, 1998). Strains positive for *eae* but not BFP were designated as atypical EPEC. Strains positive for either ETEC enterotoxins were considered ETEC and strains positive for either EAEC factor were considered EAEC. A monoplex PCR reaction targeting an accessory gene (*daaC*) of a major fimbrial sub-unit, F1845, was used for identifying DAEC (Campos et al., 1999).

Template DNA was prepared by mixing a loop-full of an overnight culture with 500µL of BP 2819 Water (Fisher Scientific, Waltham, MA, USA) in a 2 mL eppendorf tube (Eppendorf AG, Hamburg, Germany) and heating it to boiling for 20 minutes. The mixture was rapidly cooled on ice and centrifuged 5000 rpm for 2 minutes. The supernatant was collected and stored at -20°C and used as template DNA in all subsequent tests.

For the PCR reaction, 3 µL of template DNA was added to 10 µL of Quick-Load® *Taq* 2X Master Mix containing 20 mM Tris-HCl, 100 mM KCl, 3.0 mM $MgCl_2$, 0.2mM deoxynucleotide triphosphates (dNTPs) and 50 µg/mL Hot Start *Taq* DNA polymerase (New

England Biolabs). Each primer pair (0.4 µL of 20 pmol/µL) was added together with 3µL of RNase-free water to a final volume of 20 µL. PCR was performed under the following conditions: preheating at 96°C for 4 min, denaturation at 95°C for 20 secs, annealing at 57°C for 20 secs, elongation at 72°C for 1 min. PCR was performed for 35 cycles with final extension at 72°C for 7 min in an Eppendorf Mastercycler Gradient thermal cycler (Eppendorf AG).

The amplification products were separated through a 2% agarose gel and visualized by ultraviolet light trans-illumination after ethidium bromide staining. The 1-kb plus A 100-bp DNA ladder (New England BioLabs, Ipswich, MA, USA) was used as a molecular size marker in gel. Control strains employed in every PCR reaction were ETEC H10407, EAEC 042 and for EPEC strains CVD 28 (*eae* positive) and HB101 (pMAR7) (*bfpA*-positive).

For DAEC, PCR regents and consumables were as described in the multiplex assay above. Thermo-cycling conditions included: Preheating was at 95°C for 3 min, denaturation at 95°C for 30 seconds, annealing at 60°C for 30 seconds and elongation at 68°C for 2 minutes. PCR was performed for 35 cycles with a final extension at 68°C for 10 minutes. DAEC strain C1845 was used as a positive control.

2.4.1. Partial gene sequencing of PCR positive DEC isolates

PCR amplification products were separated through a 2% agarose gel, visualized by ultraviolet light trans-illumination after ethidium bromide staining, excised, and purified using a QIAquick® PCR Purification Kit (Qiagen Inc, Germantown, MD, USA) per the manufacturer's instructions. The purified PCR product was mixed with a single primer of the respective gene target (1 µL), 10 µL of DNA template and 4 µL of RNase free water to make a total of 15 µL in a PCR grade tube. The samples were delivered for final analytical

confirmation to the DNA Sequencing Service (GeneScipt USA Inc.) at The University of Virginia.

Table 1 Primer sequences and the expected amplicon sizes for the multiplex polymerase chain reaction employed in the detection of Diarrheagenic *Escherichia coli*.

Pathogen	Primer	Target Gene	Primer Sequence (5'-3')	Amplicon (bp)	
	LT-F	*elt*	CACACGGAGCTCCTCAGTC	508	173
	LT-R		CCCCCAGCCTAGCTTAGTTT		
	ST-F	*est*	GCTAAACCAGTAG/AGGTCTTCAAAA	147	174
	ST-R		CCCGGTACAG/AGCAGGATTACAACA		
EPEC	BFPA-F	*bfpA*	GGAAGTCAAATTCATGGGGG	367	
	BFPA-R		GGAATCAGACGCAGACTGGT		175
	EAE-F	*eae*	CCCGAATTCGGCACAAGCATAAGC	881	
	EAE-R		CCCGGATCCGTCTCGCCAGTATTCG		176
EAEC	CVD432F	*aatA*	CTGGCGAAAGACTGTATCAT	630	
	CVD432R		CAATGTATAGAAATCCGCTGTT		
	AAIC F	*aaiC*	ATTGTCCTCAGGCATTTCAC	215	177
	AAIC R		ACGACACCCCTGATAAACAA		
DAEC	DAAC F	*daaC*	ATTACGTCATCCGGGAAGCACACA	146	178
	DAAC R		TTGTCTGCGGCTTTATGAGCAAGC		

Abbreviations: EAEC, enteroaggregative *Escherichia coli*; EPEC, enteropathogenic *E. coli*; ETEC, enterotoxigenic *E. coli*; EAEC, EPEC and ETEC were dentified using a multiplex PCR and primer pairs according to (Panchalingam et al., 2012) and DAEC using a monoplex PCR and primer pairs according to Campos et al., 1999)

2.5. Cell adherence assay

As HEp-2 or HeLa cell adherence remains the gold standard for discrimination of DEC pathotypes showing localized (EPEC), aggregative (EAEC) and diffuse (DAEC) adherence (Croxen et al., 2013; Kaper, 2005; Nataro and Kaper, 1998), 46 out of the total 205 strains were characterized. The Hep-2 cell adherence assay as described by (Nataro et al., 1987) for differentiating patterns of DEC was used with some modifications. HeLa cells instead of Hep-2 cells were used to determine the adherence patterns of 46 out of the 205 *E. coli* strains from that showed poor resolution with PCR using the *daaC* gene (DAEC). However, all previously characterized EAEC strains based on PCR and partial gene sequencing of virulence gene determinants were excluded from this analysis. Selection of these strains was based on poor resolution of presumptive DAEC identified using PCR. HeLa cells at 80% confluence were aseptically transferred into 24-well plates (Fisher Scientific) containing 12 mm cover slips (Fisher Scientific) in each well and washed with Phosphate Buffer Saline (PBS) (Fisher Scientific) and 1 mL of Dulbecco's Modified Eagle's Medium (DMEM) (Sigma-Aldrich) +0.5% D-mannose (Sigma-Aldrich). The bacterial isolates were grown in 3 mL of LB broth in 13 mL plastic tubes while shaking at 37°C. Ten millilitres of the overnight culture was added to each well using standard strains EPEC E2348/69, EAEC 042 and DAEC 1845 as controls of localized, aggregative and diffuse adherence, respectively. Incubation was at 37°C in a CO_2 incubator for 3 and 6 hours (on replenishing with fresh medium after 3 h) for each isolate. The cells were washed gently (3X) with Peptone Buffered Saline (PBS) (Sigma-Aldrich) and 500 µL of 2% formalin were added to fix the samples for 20 min at room temperature. The samples were rinsed (3X) with distilled water (dH_20) and stained with 500 µL of a Giemsa solution (Sigma-Aldrich) for 20 minutes. The samples were rinsed (3X) with dH_20 until the colour disappeared. The coverslips were removed from the 24-well plates, airdried and mounted with a tiny drop of Cytoseal (Fisher Scientific) mounting glue onto a

glass slide. The samples were observed under a Zeiss light microscope and images were recorded at 60 X.

2.6. Statistical analysis

Means for prevalence were calculated for each identified DEC pathotype based on the method of analysis (molecular or cell adherence) for all strains isolated from food sources and irrigation water. We used the Basic Local Alignment Search Tool (BLAST) within National Center for Biotechnology Information (NCBI) to search for nucleotide sequences of *E. coli* strains deposited in GenBank, with ≥80% similarity to *aaiC* or *aatA* sequences found in EAEC strains in this study. Thereafter, the evolutionary relationship of each strain was inferred using the Neighbour-Joining method and computed with Molecular Evolutionary Genetics Analysis for Bigger Data Sets version 7.0 (MEGA 7) (Kumar et al., 2016).

3. Results

3.1. Haemolysin production

Of the 205 isolates tested on blood agar, only 1 strain from irrigation water (CR4) showed beta-haemolysis and none showed alpha-haemolysis. As haemolysis is correlated with the presence of uropathogenic *E. coli*, we concluded that these strains were not common in our samples.

3.2. PCR and partial virulence gene sequencing

Among the different DEC virulence gene determinants sought in our collection of 205 *E. coli* strains, only those specific for EAEC and DAEC were found by PCR. The remaining 196 isolates were negative for virulence gene determinants of the DEC pathotypes screened for in this study using PCR. Based on PCR (3.9%, 8 out of 205) and (0.5%, 1 out of 205) were EAEC and DAEC respectively (Table 2). EAEC was predominant among strains from PDBM (5.1%, 6 out of 118) and to a lesser extent irrigation water (4.2%, 2 out of 48) of which a single isolate (K2) carried the *aaiC* gene and 5 isolates (M1, 57, L7, N25 and 79) carried the *aatA* gene. The single strain (CR4) of DAEC was from irrigation water.

Partial gene sequencing of individual virulence gene determinants associated with each pathotype among PCR positive strains (excluding strain 79) confirmed only EAEC.

Gel electrophoresis bands showing strain 79 after PCR consistently showed a faint signal compared to the other sequenced strains (data not shown) and hence was left out of the sequencing analysis. However, numbers of EAEC were lower compared to those observed based solely on characterization with PCR (Table 2) suggesting an initial over estimation of EAEC prevalence. Based on partial gene sequencing, EAEC was found in 2.4%, 5 out of 205 strains. EAEC strains confirmed with partial gene sequencing included K2 positive for *aaiC* (Table 3).

On the other hand, strains 57, L7, MPU(W)5(1) and MPU(W)8(4) were positive for *aatA* (Table 3). EAEC was confirmed in PDBM (2.5%, 3 out of 118) and irrigation water (4.2%, 2 out of 48).

3.3. Evolutionary relationship of EAEC strains isolated from food sources and irrigation in South Africa with genetically related *E. coli* strains in GenBank based on partial gene sequencing of *aaiC* and *aatA* genes

We used the evolutionary relationship of EAEC strains isolated from this study to infer relatedness to *E. coli* strains previously isolated from different sources and deposited within GenBank as a way of determining potential routes for DEC contamination. The nucleotide sequence of the *aaiC* gene in strain K2 from PDBM showed 80% identity to 15 *E. coli* strains in GenBank. The dendrogram constructed to infer the evolutionary relationship of strain K2 with these 15 strains generated 4 distinct clusters; G1, G2, G3 and G4. The strains predominantly clustered based on location of *aaiC* within the bacterium (chromosome or plasmid) well as on the source of strain isolation (food sources or human stool) (Fig. 1). G1 predominantly consisted of strains of *E. coli* serovar O104:H4 among which were outbreak strains 227-11 and 2011C-3493 isolated from patients in Germany and The United States respectively. The closest strain related to K2 was also isolated from a food source, *E. coli* strain 06-0048 (Accession number: CP012498.1) isolated from alfalfa sprouts in California USA in 2006. Strains isolated from humans (G1 and G4) were isolated from bloody and non-bloody stool in patients from Denmark, France, Georgia, Poland and the USA suggesting a link of the *aaiC* gene with diarrheal causing *E. coli*.

The nucleotide sequence of the *aatA* gene in strain 57 from milk showed between 90 to 93% identity to 22 *E. coli* strains in GenBank (Fig. 3). However, nucleotide sequences from strains L7 isolated from milk as well as MPUW51 and MPU84 both isolated from irrigation water did

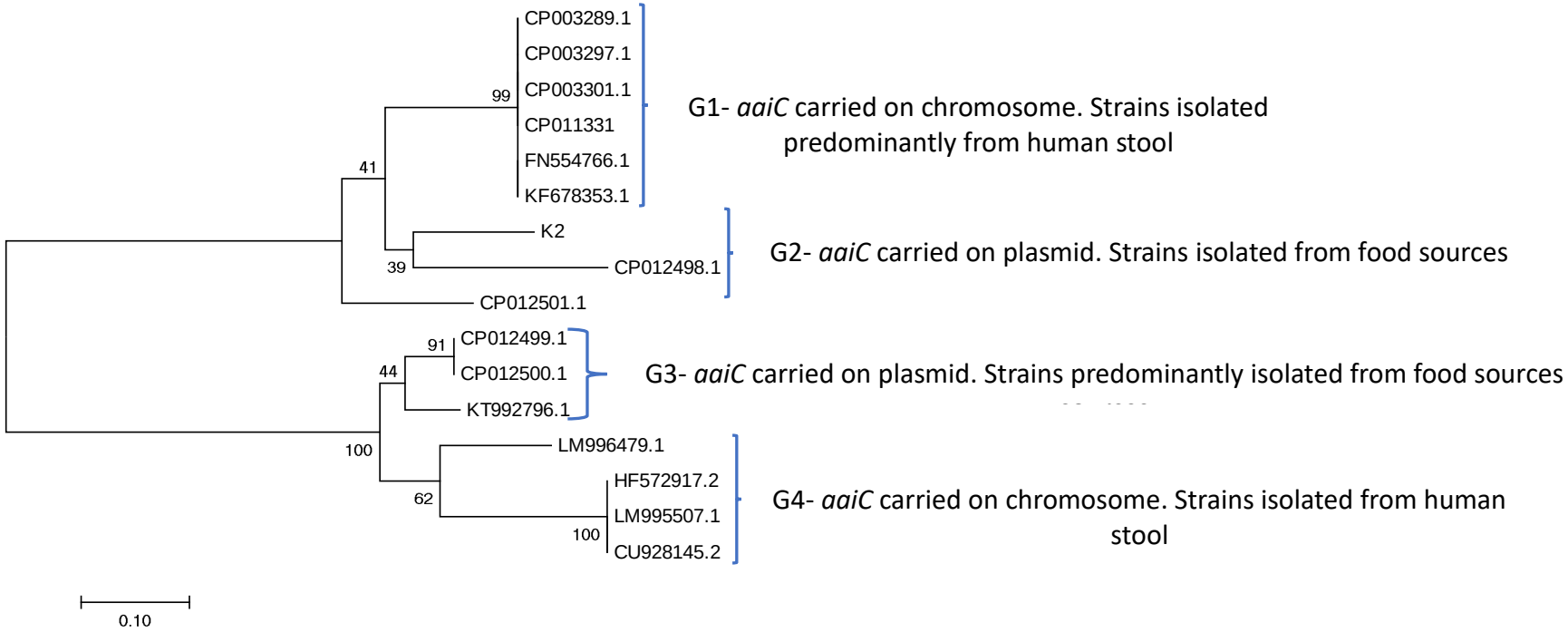

Fig. 1.Evolutionary relationships of *Escherichia coli* strains in GenBank showing ≥80% gene nucleotide sequence similarity to strain K2 isolated from producer distributor bulk milk in South African milk based on aaiC, the enteroaggregative *Escherichia coli* (EAEC) virulence gene determinant. The evolutionary history was inferred using the Neighbour-Joining Method. Codes of strains used for comparison represent accession numbers from GenBank. G1, G2, G3 and G4 represent defined clusters of strains showing difference in ge netic location of aaiC (plasmid or chromosome) in each strain as well as source of strain isolation (foodborne or human faeces).

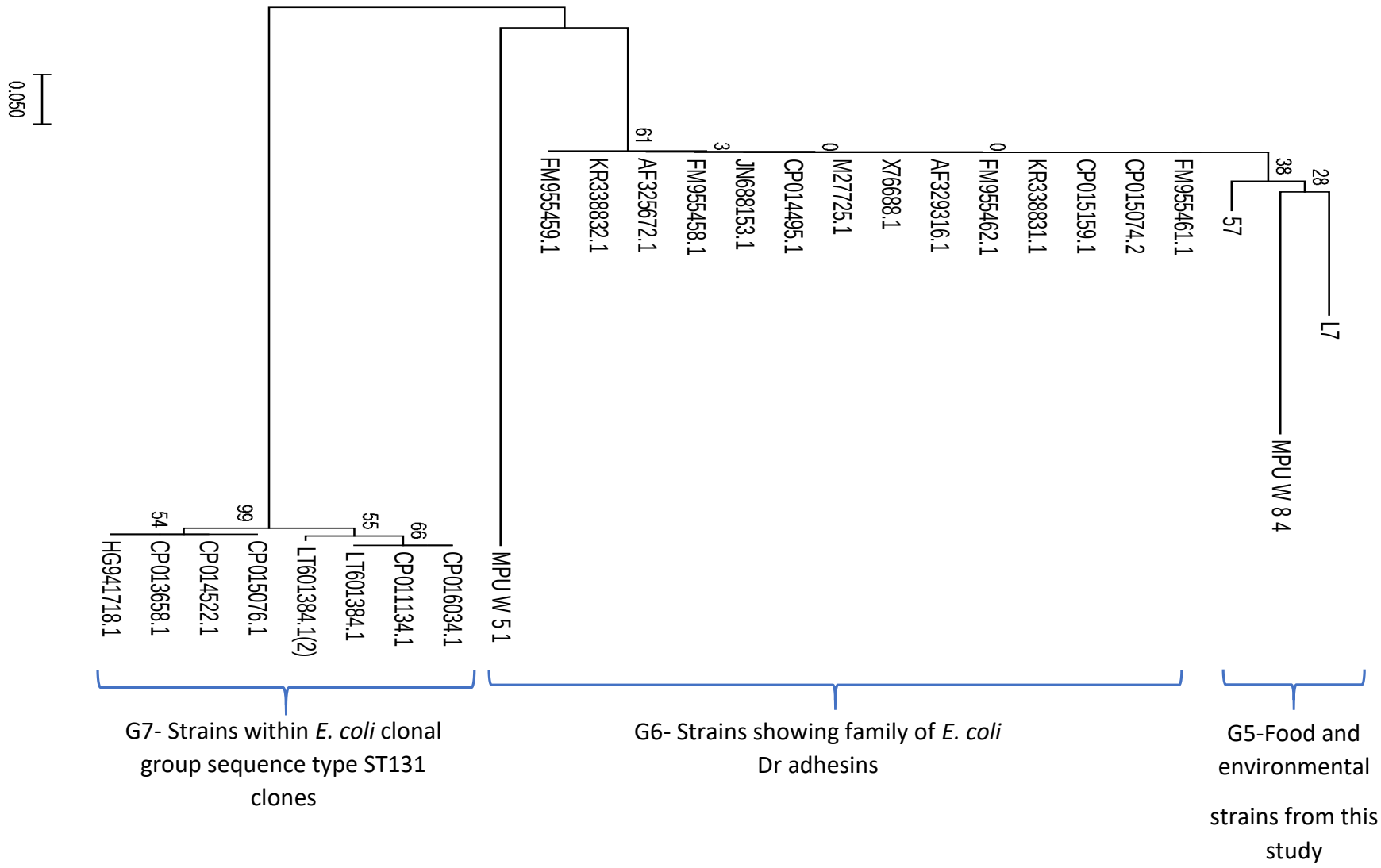

Fig. 2. Evolutionary relationships of *Escherichia coli* in GenBank showing ≥90% gene nucleotide sequence similarity to *E. coli* strains isolated from producer distributor bulk milk (PDBM) and irrigation water in South Africa based on *aatA*, the enteroaggregative *E. coli* (EAEC) virulence gene determinant. The evolutionary history was inferred using the Neighbour-Joining Method. Strain sources: PDBM-L7, 57; Irrigation water-MPUW51, MPUW84. Codes of strains used for comparison represent accession numbers in GenBank.

not have comparisons in GenBank. Nonetheless, we used the 22 strains showing similarity to the single strain 57 for drawing evolutionary relationships with all 4 strains. The dendrogram inferring evolutionary relationships generated 3 clusters G5, G6 and G7 (Fig. 2).

G5 comprised 3 of the 4 strains (PDBM=2; irrigation water=1) used for the analysis. Inspite of clustering together, strains in G5 showed low evolutionary relatedness (Fig. 2). G6 contained 1 isolate from irrigation water (MPUW51) that showed closest relatedness to *E. coli* strains positive for the *Dr* family of adhesins. Additionally, most isolates in G5 and G6 were isolated from humans. G7 consisted of strains predominantly within the clonal group, sequence type ST131. Strains from G6 and G7 were predominantly isolated from clinical sources. Similarly, as previously noted with strains clustered based on the *aaiC* gene, strains from this study predominantly clustered based on source of isolation (environmental or clinical). Strains isolated from humans (G6 and G7) were predominantly associated with extraintestinal infections such as urinary tract infections in patients from China, France, Poland, Spain and the UK suggesting a link of *aatA* with extraintestinal pathogenic *E. coli*.

3.4 Adherence tests

The strains predominantly exhibited the characteristic ‘stacked-brick’ pattern or aggregative adherence (AA) of EAEC and to a less extent invasiveness typical of EIEC on HeLa cells (Table 2). No strain showed the diffuse adherence phenotype typical of DAEC suggesting false positives with PCR that led to the poor resolution previously reported. Based on the adherence and invasive phenotypes, EAEC and EIEC were found in (24%, 11 out of 46) and (4%, 2 out of 46) of strains respectively. EAEC strains exhibited strong to moderate and weak AA capacity (Fig. 3a, 3b and 3c). Prevalence of EAEC in PDBM, irrigation water and irrigated lettuce was (15%, 7 out of 46), (7%, 3 out of 46) and (2%, 1 out of 46) respectively (Table 2). Prevalence of EIEC was 2%, 1 out of 46 in both PDBM and irrigated lettuce.

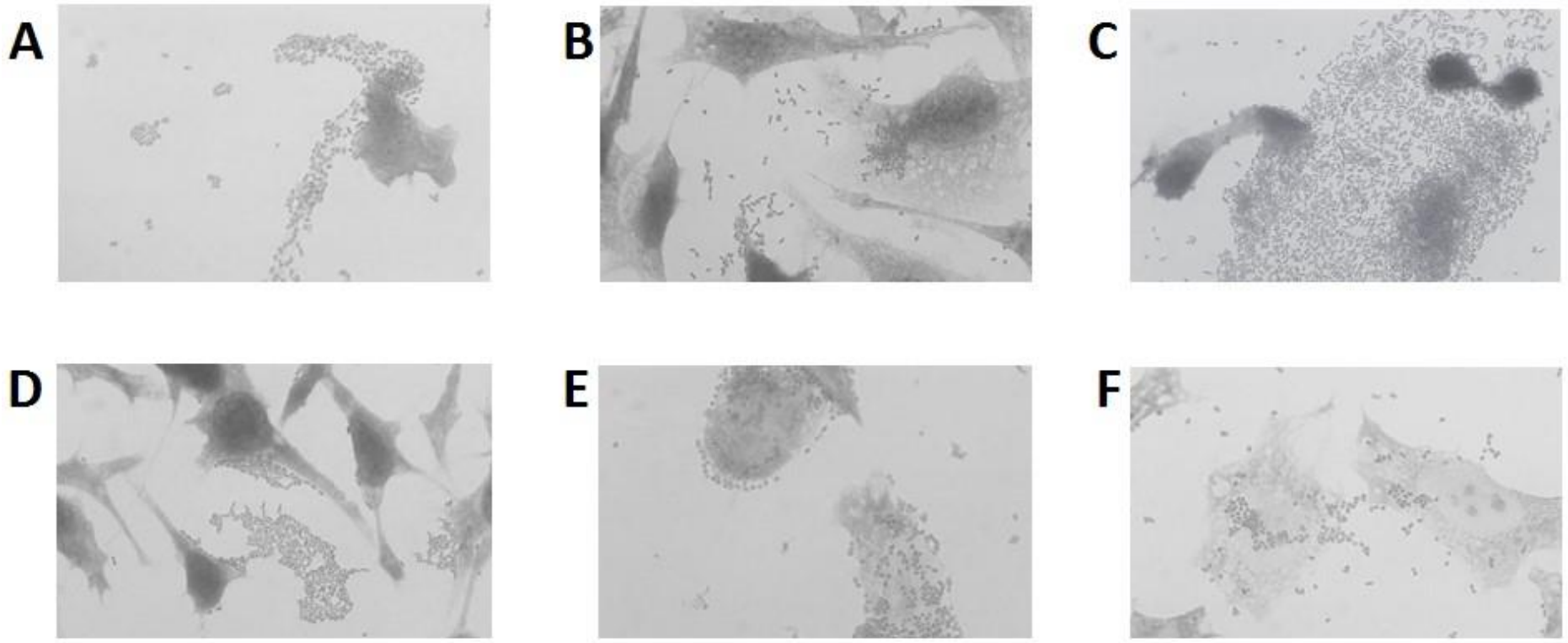

Fig. 3a. Aggregative adherence (AA) pattern "stacked-brick" observed in *Escherichia coli* strains isolated from food sources and irrigation water in South Africa. Strains were grown on HeLa cells and showed strong to moderate AA characteristic of enteroaggregative *Escherichia coli* (EAEC). Strain code: Standard EAEC strain 042; B-H8; C-M28; D-NW(V)7(3); E-NW(V)10(1); F-NW(W)9(3). Source of isolation: A-Clinical strain; B and C-Producer distributor bulk milk; D and E-Irrigated lettuce; F-Irrigation water. Images were taken to 100× with a Zeiss microscope. Resolution = 20 μm..

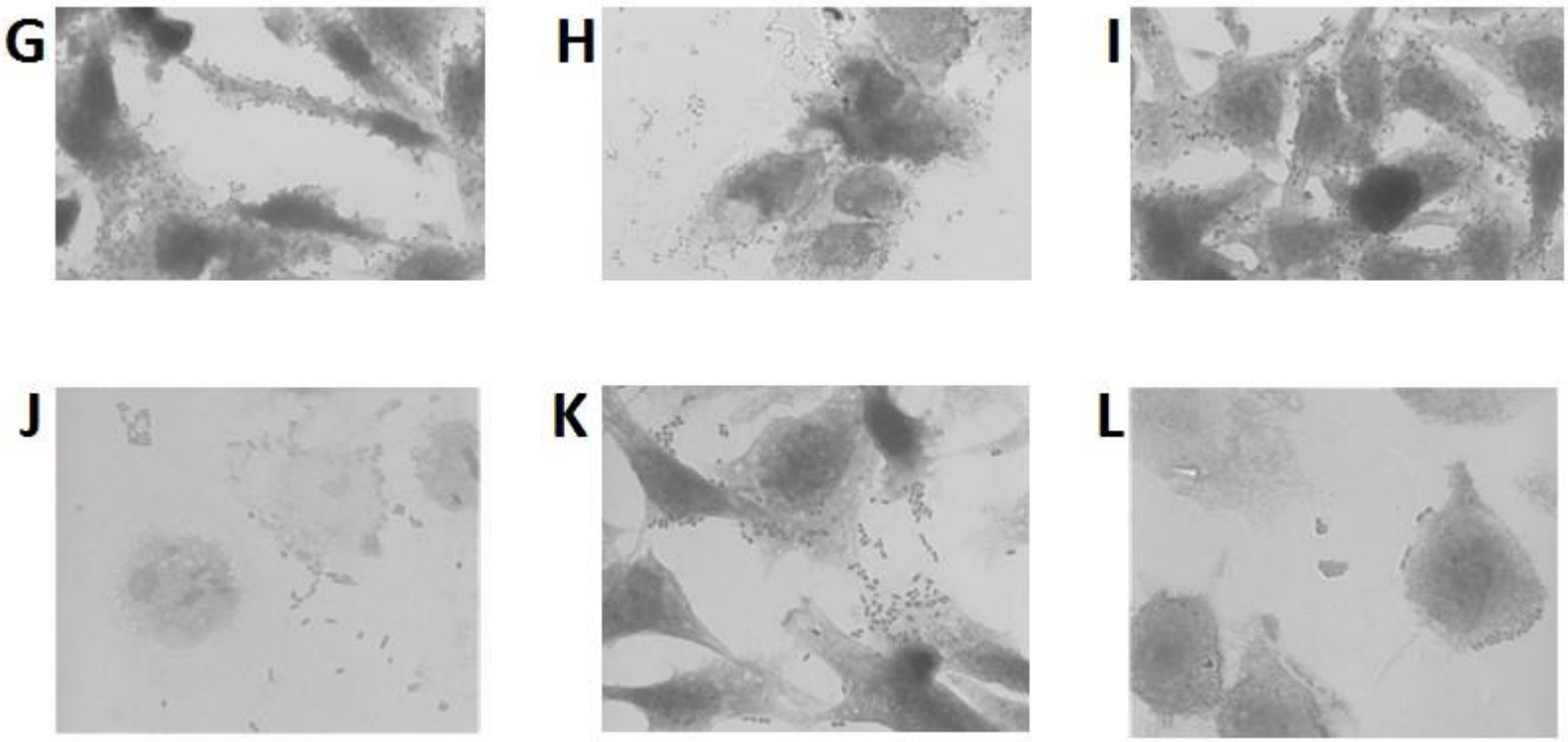

Fig. 3b. Weak aggregative adherence (AA) observed in *Escherichia coli* strains isolated from food sources and irrigation water in South Africa. Strains were grown on HeLa cells and showed weak adherence in comparison to strains in Fig. 3a. Strain codes: G- LeK1; H-M12; I-K5; J-K16; K-M6; L-N23. Strain sources: G-Irrigated lettuce; H, I, J, K and L-Producer distributor bulk milk. Images were taken to 100× with a Zeiss microscope. Resolution=20µm.

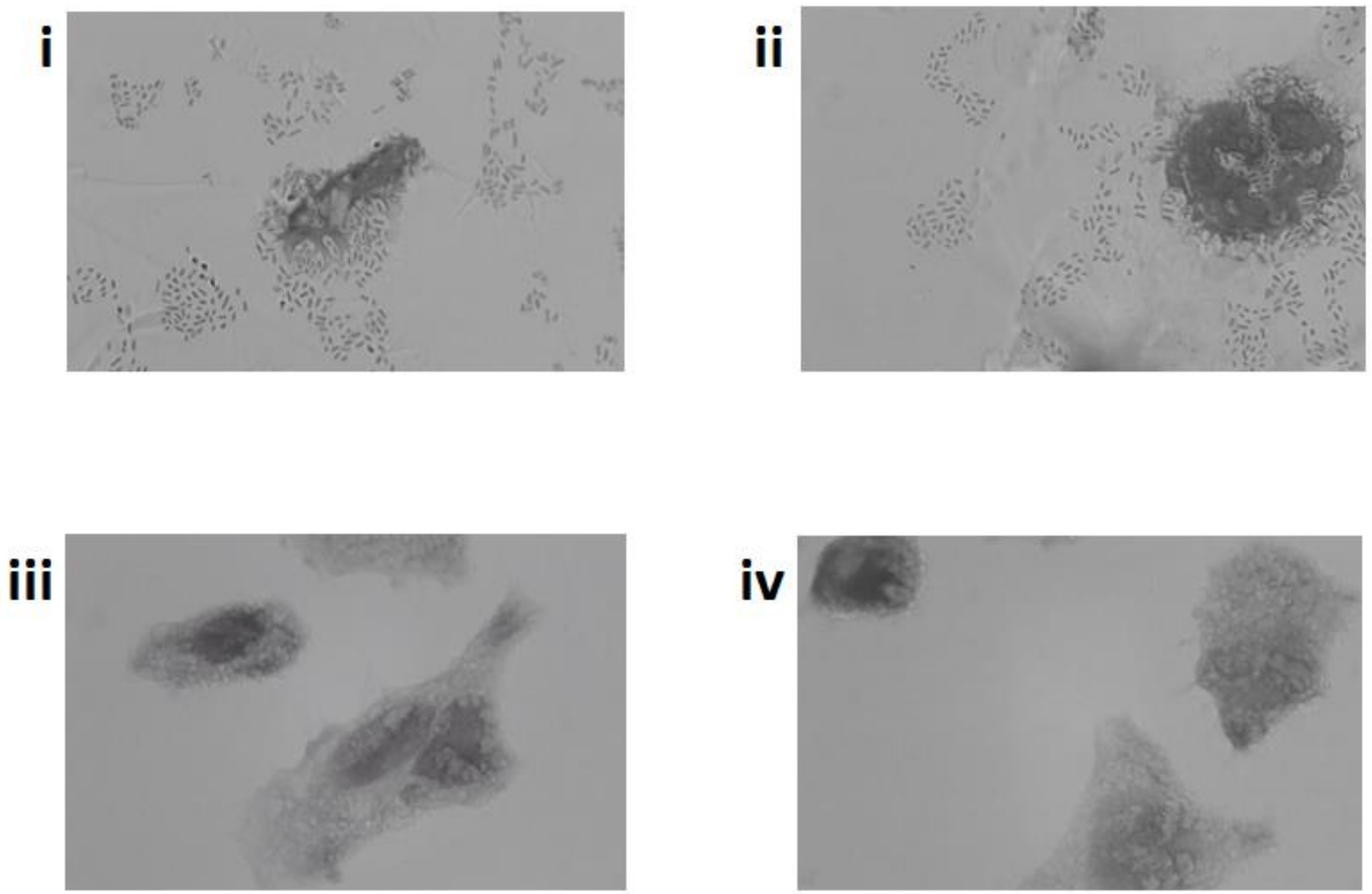

Fig.3c. Invasiveness observed in *Escherichia coli s*trains isolated from food sources and irrigation water in South Africa. Invasivenes is characteristic of enteroinvasive *Escherichia coli* (EIEC) and therefore strains were identified as presumptive EIEC since no other subsequent confirmatory test was done. Strains were grown on HeLa cells. Strain codes: i and ii-LeK; iii and iv-37. Strain sources: LeK-Irrigated lettuce; 37-Producer distributor bulk milk. Images were taken at 60× with a Zeiss microscope. Resolution=20µm.

Table 2 Percentage of *Escherichia coli* strains isolated from food sources and irrigation water in South Africa positive for virulence gene determinants(n=205) and subsequently cell adherence patterns (n=46) associated with Diarrheagenic *E. coli*.

Sample	Method of characterization							
	Polymerase chain reaction				Partial gene sequencing		HeLa cell adherence pattern	
	Total number of isolates	*aaiC* (EAEC)	*aatA* (EAEC)	*da* (DAEC)	*aaiC* (EAEC)	*aatA* (EAEC)	Aggregative adherence (EAEC)	Cell invasiveness (EIEC)
PDBM	118	1(1)	4(5)	-	1(1)	2(2)	15(7)	2(1)
Irrigation water	48	-	4(2)	2(1)	-	4(2)	7(3)	-
Irrigated lettuce	29	-	-	-	-	-	2(1)	2(1)
Coleslaw	10	-	-	-	-	-	-	-
Total	205	1	7	1	1	4	24(11)	2(2)

Percentages reported to the first significant whole number. Number of isolates in parentheses; - not detected/no adherence pattern; *aaiC*-Chromosomal virulence gene determinant in EAEC; *aatA*- Plasmid virulence gene determinant in EAEC; EAEC-enteroaggregative *E. coli*;

daaC- a major fimbrial sub-unit in standard DAEC strain C1845; DAEC-diffusely adherent *E. coli*; EAEC-enteroaggregative *Escherichia coli*; EIEC-enteroinvasive *E. coli*. All strains tested for cell adherence had previously shown poor resolution using PCR.

Table 3 Nucleotide sequences of virulence gene determinants associated with enteroaggregative *Escherichia coli* (EAEC) in *Escherichia coli* strains isolated from food sources and irrigation water in South Africa.

Isolate code	Isolate source	Sequenced gene product (*aaiC* or *aatA*)
K2	PDBM	CTTGGACGGCCTGTTTTTGATTTGGTTATATTAAGTTAAAGGTCACTAAGGAAATACAACTCTTATTA GAGCATTTTTTTAAAAGAGGTGAGAAAGATATTGTAATTGAAATTGTTCGCAAGGAATCTACTAAAT CAGGTAGTGCATATTCATCCTTTAAGGTTGTTTATCAGGGGTGTCGTAA
57	PDBM	TAAACCGCCTGGGTCAGCGCATCGTGGGTGGGGAACACCGGGCTGAACACGGGGCCACCGGACTGG CAGACCGTGGAATAAGGCATCGGGGGTGAACGTTGTCTGCGGCTTTATGAGCAAGCAGCCCTT
L7	PDBM	TAGCTGTAGCGCATCGTGAGGGCGA ACCCGGCTGACCCGGGCCACCGGAC TGCGTACGTGTAAAACAGACAGCGT GCGCT GAGCGTTGTCTGCGCTTTATGAGCA GCTCGCTCTTTTAGAGGG
MPU(W)5(1)	Irrigation water	ATTACTCTTTTCCATTTGCATCCCCT TGGCTAAACCAGCTTACTCTATCACC GGACTGGCCGATAGTTTATAATCAA TCC GTCTGGTTTTAACAGAGTTATGAGA ATCAGTTCCATTTTATTTGGCCTTGG TTTCTCTTGATTTCTGGTGAAA
MPU(W)8(4)	Irrigation water	ACTGGCGTCCCGCCGCATCGTGTTTG GGCGAACACCGGGCACAGCCGGCGC CACCGGATGGGGACCGGGAAATTAC GAAT GGGCGTGAGCGTTGGCGGCGGCTTT GGAGAGGCAGCCCCTTTCGAATAAA G

PDBM-Producer Distributor Bulk Milk; K2-positive for *aaiC*; 57, L7, MPU(W)5(1) and MPU(W)8(4)-positive for *aatA*

4. Discussion

Determining the most prevalent DEC pathotypes associated with diarrheal disease within a given locale can help provide better guidance to food safety and public health interventions. Predominance of EAEC in food sources and irrigation water in South Africa suggests it is the most prevalent DEC pathotype in these sources. These sources subsequently provide routes for EAEC infection within the general population. Therefore, EAEC may be the leading cause of food and water-borne diarrheal disease caused by pathogenic *E. coli* in South Africa. EAEC is an emerging food-borne pathogen largely affecting infants, immune compromised adults and travellers to developing countries (Croxen et al., 2013; Estrada-Garcia and Navarro-Garcia, 2012). Additionally, EAEC has been associated with inflammation and malnutrition among infants in developing countries (Acosta et al., 2016) thereby heightening the risk of a similar scenario playing out locally (South Africa) especially within the low income urban and rural population who may lack adequate food and water safety systems. Previous studies in South Africa have also noted high prevalence of EAEC in food and environmental (Caine et al., 2014) as well as clinical sources (Adefisoye and Okoh, 2016; Bisi-johnson et al., 2011; Samie et al., 2007; Tau et al., 2012).

Aaic forms part of a group of genes localized to a 117kb pathogenicity island inserted at pheU in typical EAEC having homologues in other Gram-negative bacteria and proposed to constitute a Type VI secretion system (T6S) (Dudley et al. 2006). Bacteria have evolved different regulatory components for the expression of these genes which may be acquired by horizontal transfer for specific adaptation to varying hosts and niches such as marine, plants or animals (Boyer et al. 2009) an attribute that was noticed among strains from our work (Fig 1). The regulatory mechanism governing type VI gene expression vary widely from species to species and even from strain to strain within a given species given its wide distribution among different bacterial groups (Miyata et al. 2013).

Additionally, the aggregative adherence in EAEC is a bacterial outcome of co-evolution of human hosts and has been acquired by different *E. coli* lineages some of which may share similar but non-identical mobile elements and so no single strain can be considered representative of EAEC (Okeke et al. 2010).

The *aatA* gene in EAEC forms part of a plasmid encoded locus also under the control of *aggR* coding for an ABC transporter complex that channels the virulence factor dispersin out of the bacterial cell (Nishi et al. 2003). On the other hand the Dr- family of adhesins consists of clones such as *Dr*, *Afa-I*, *Afa-III* and *F1845* commonly associated with Uropathogenic *E. coli* (UPEC) but also found in DEC. Incidentally *agg* and *aaf* adhesins of standard EAEC strains 17-2 and 042 respectively have identity to *afaD-III* an afimbrial adhesin in UPEC and DEC strains (Garcia et al., 2000, 1996). While the specific role of the *aatA* in EAEC and Dr family of adhesins in pathogenic *E. coli* may differ, both are involved in colonization of the host environments and thereby persistence.

Predominance of EAEC within liquid environments (milk and irrigation water) compared to solids (irrigated lettuce and coleslaw) suggests provision of more favourable conditions for persistence, thereby presenting higher risk as potential routes for foodborne infection. Additionally, EAEC form biofilms on abiotic surfaces (Estrada-Garcia and Navarro-Garcia, 2012), an adaptation that may facilitate persistence within a secondary environment providing a protection against external stresses as well as capturing nutrients. This adaptation poses a risk factor for contamination of food contact surfaces.

The low prevalence of potentially invasive *E. coli* (presumptive EIEC) and absence of other DEC pathotypes suggests that food sources and irrigation water in South Africa may not be major reservoirs/carriers of these pathotypes. Additionally, these absent pathotypes may not serve as causes of DEC associated illness in South Africa.

Therefore, compared to other DEC pathotypes, EAEC strains maybe more suitably adapted for longer persistence within the open environment as has previously been noted with outbreak strains of enteroaggregative haemorrhagic *E. coli* (EAHEC) serotype O104:H4.

Prevalence of DEC within this study may have been underestimated based characterization of strains using molecular tools (PCR and sequencing). This is because subsequent characterization of poorly resolved strains (PCR) with cell adherence assays identified more DEC pathotypes (EAEC and invasive *E. coli* (presumptive EIEC)). This observation highlights shortcomings associated with molecular diagnostic tools targeting virulence gene determinants during monitoring and surveillance (due to the impracticality of tissue culture assays) studies especially when targeting pathogens with great heterogeneity such as EAEC. Therefore, studies investigating the prevalence of DEC within food and environmental sources maybe underestimating the levels and subsequently the risk posed to food safety and public health.

However, on a brighter note, partial gene sequencing of virulence gene determinants followed by inference of evolutionary relationships seems to provide an adequate and cost-effective means of source tracking foodborne, environmental and clinical DEC. We show that using a single sequenced virulence gene coupled with its comparison to strains of the same species within GenBank having the same gene with high nucleotide sequence similarity (≥80%) provides a quick and accurate tracker for contamination sources. This approach unlike sophisticated and expensive techniques such as Pulse Field Gel Electrophoresis (PFGE) and Whole Genome Sequencing (WGS) may be more applicable within resource limited areas.

Lastly, the large diarrheal disease burden noticed in South Africa especially among infants (Chola et al., 2010) and immune compromised adults such as HIV-positive individuals (Moshabela et al., 2012) may be associated with EAEC transmitted through contaminated

food and water. Therefore, routine screening for this pathogen in food and environmental sources especially among low income communities may help monitor and control risks of potential outbreaks and long-term health and nutritional effects emanating from recurrent infections.

Conflict of interest

The authors declare no conflict of interest

Acknowledgements

The Department of Research and Innovation, University of Pretoria for a post-graduate travel bursary to Matthew Aijuka to travel to James P. Nataro's laboratory at The University of Virginia. Work in the Nataro lab was supported by US National Institutes of Health grant AI-33096 to JPN.

Staff and members of Dr. Nataro and Dr. Girón laboratories for technical and laboratory assistance.

Aquillah M. Kanzi from Forestry and Agricultural Biotechnology Institute (FABI), Department of Genetics, University of Pretoria for assistance with phylogenetic analysis

References

Acosta, G.J., Vigo, N.I., Durand, D., Riveros, M., Arango, S., Zambruni, M., Ochoa, T.J., 2016. Diarrheagenic Escherichia coli: Prevalence and Pathotype Distribution in Children from Peruvian Rural Communities. Am. J. Trop. Med. Hyg. 95, 574–579. doi:10.4269/ajtmh.16-0220

Adefisoye, M.A., Okoh, A.I., 2016. Identification and antimicrobial resistance prevalence of pathogenic Escherichia coli strains from treated wastewater effluents in Eastern Cape, South Africa. Microbiologyopen 5, 143–151. doi:10.1002/mbo3.319

Aijuka, M., Charimba, G., Hugo, C.J., Buys, E.M., 2015. Characterization of bacterial pathogens in rural and urban irrigation water. J. Water Health 13, 103–117. doi:10.2166/wh.2014.228

Aijuka, M.E.O., 2014. Bacteriological quality of South African irrigation water and its role as a source of contamination on irrigated lettuce. University of Pretoria, Pretoria.

Baker, K.K., O'Reilly, C.E., Levine, M.M., Kotloff, K.L., Nataro, J.P., Ayers, T.L., Farag, T.H., Nasrin, D., Blackwelder, W.C., Wu, Y., Alonso, P.L., Breiman, R.F., Omore, R., Faruque, A.S.G., Das, S.K., Ahmed, S., Saha, D., Sow, S.O., Sur, D., Zaidi, A.K.M., Quadri, F., Mintz, E.D., 2016. Sanitation and Hygiene-Specific Risk Factors for Moderate-to-Severe Diarrhea in Young Children in the Global Enteric Multicenter Study, 2007-2011: Case-Control Study. PLoS Med. 13, e1002010. doi:10.1371/journal.pmed.1002010

Bisi-johnson, M.A., Obi, C.L., Vasaikar, S.D., Baba, K.A., Hattori, T., 2011. Molecular basis of virulence in clinical isolates of Escherichia coli and Salmonella species from a tertiary hospital in the Eastern Cape , South Africa 1–8. doi:10.1186/1757-4749-3-9

Caine, L.-A., Nwodo, U., Okoh, A., Ndip, R., Green, E., 2014. Occurrence of Virulence Genes Associated with Diarrheagenic Escherichia coli Isolated from Raw Cow's Milk from Two Commercial Dairy Farms in the Eastern Cape Province, South Africa. Int. J. Environ. Res. Public Health 11, 11950–11963. doi:10.3390/ijerph111111950

Campos, L.C., Vieira, M.A., Trabulsi, L.R., da Silva, L.A., Monteiro-Neto, V., Gomes, T.A., 1999. Diffusely adhering Escherichia coli (DAEC) strains of fecal origin rarely express F1845 adhesin. Microbiol. Immunol. 43, 167–170.

Carlton, E.J., Woster, A.P., DeWitt, P., Goldstein, R.S., Levy, K., 2016. A systematic review and meta-analysis of ambient temperature and diarrhoeal diseases. Int. J. Epidemiol. 45, 117–130. doi:10.1093/ije/dyv296

Castro-Rosas, J., Cerna-Cortés, J.F., Méndez-Reyes, E., Lopez-Hernandez, D., Gómez-Aldapa, C.A., Estrada-Garcia, T., 2012. Presence of faecal coliforms, Escherichia coli and diarrheagenic E. coli pathotypes in ready-to-eat salads, from an area where crops are irrigated with untreated sewage water, International Journal of Food Microbiology. doi:10.1016/j.ijfoodmicro.2012.03.025

Chola, L., Michalow, J., Tugendhaft, A., Hofman, K., 2010. Reducing diarrhoea deaths in South Africa: costs and effects of scaling up essential interventions to prevent and treat diarrhoea in under-five children. doi:10.1186/s12889-015-1689-2

Croxen, M.A., Law, R.J., Scholz, R., Keeney, K.M., Wlodarska, M., Finlay, B.B., 2013. Recent advances in understanding enteric pathogenic Escherichia coli. Clin. Microbiol. Rev. 26, 822–880. doi:10.1128/CMR.00022-13

Estrada-Garcia, T., Navarro-Garcia, F., 2012. Enteroaggregative Escherichia coli pathotype: A genetically heterogeneous emerging foodborne enteropathogen. FEMS Immunol. Med. Microbiol. 66, 281–298. doi:10.1111/j.1574-695X.2012.01008.x

Farrokh, C., Jordan, K., Auvray, F., Glass, K., Oppegaard, H., Raynaud, S., Thevenot, D., Condron, R., De Reu, K., Govaris, A., Heggum, K., Heyndrickx, M., Hummerjohann, J., Lindsay, D., Miszczycha, S., Moussiegt, S., Verstraete, K., Cerf, O., 2013. Review of Shiga-toxin-producing Escherichia coli (STEC) and their significance in dairy production. Int. J. Food Microbiol. 162, 190–212. doi:10.1016/j.ijfoodmicro.2012.08.008

Garcia, M.-I., Gounon, P., Courcoux, P., Labigne, A., Le Bouguénec, C., 1996. The afimbrial adhesive sheath encoded by the afa-3 gene cluster of pathogenic Escherichia coli is composed of two adhesins. Mol. Microbiol. 19, 683–693. doi:10.1046/j.1365-2958.1996.394935.x

Garcia, M.-I., Jouve, M., Nataro, J.P., Gounon, P., Le Bouguénec, C., 2000. Characterization of the AfaD-like family of invasins encoded by pathogenic *Escherichia coli* associated with intestinal and extra-intestinal infections. FEBS Lett. 479, 111–117. doi:10.1016/S0014-5793(00)01898-6

Gemmell, M.E., Schmidt, S., 2012. Microbiological assessment of river water used for the irrigation of fresh produce in a sub-urban community in Sobantu, South Africa. Food Res. Int. 47, 300–305. doi:10.1016/j.foodres.2011.07.016

Greene, N.P., Crow, A., Hughes, C., Koronakis, V., 2015. Structure of a bacterial toxin-activating acyltransferase. Proc. Natl. Acad. Sci. U. S. A. 112, E3058-66. doi:10.1073/pnas.1503832112

Kaper, J.B., 2005. Pathogenic Escherichia coli. Int. J. Med. Microbiol. 295, 355–356. doi:10.1038/nrmicro818

Kumar, S., Stecher, G., Tamura, K., 2016. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. Mol. Biol. Evol. 33, msw054.

doi:10.1093/molbev/msw054

Moshabela, M., MacPherson, P., Ezard, N., Frean, E., Mashimbye, L., Elliott, J.H., Oldenburg, B., 2012. Clinical and social determinants of diarrhoeal disease in a rural HIV/AIDS clinic, South Africa: a case-control study. Int. J. STD AIDS 23, 346–350. doi:10.1258/ijsa.2011.011285

Nash, J.H., Villegas, A., Kropinski, A.M., Aguilar-Valenzuela, R., Konczy, P., Mascarenhas, M., Ziebell, K., Torres, A.G., Karmali, M. a, Coombes, B.K., 2010. Genome sequence of adherent-invasive Escherichia coli and comparative genomic analysis with other E. coli pathotypes. BMC Genomics 11, 667. doi:10.1186/1471-2164-11-667

Nataro, J., Kaper, J., 1998. Diarrheagenic Escherichia coli Strains. Clin. Mirobiology Rev. 11, 142–201.

Nataro, J.P., Kaper, J.B., Robins-Browne, R., Prado, V., Vial, P., Levine, M.M., 1987. Patterns of adherence of diarrheagenic escherichia coli to HEp-2 cells. Pediatr. Infect. Dis. J. doi:10.1097/00006454-198709000-00008

Newell, D.G., Koopmans, M., Verhoef, L., Duizer, E., Aidara-Kane, A., Sprong, H., Opsteegh, M., Langelaar, M., Threfall, J., Scheutz, F., der Giessen, J. van, Kruse, H., 2010. Food-borne diseases — The challenges of 20years ago still persist while new ones continue to emerge. Int. J. Food Microbiol. 139, S3–S15. doi:10.1016/j.ijfoodmicro.2010.01.021

Ntuli, V., Njage, P.M.K., Buys, E.M., 2017. Extended-spectrum β-lactamase, shigatoxin and haemolysis capacity of O157 and non-O157 E. coli serotypes from producer-distributor bulk milk. doi:10.1016/j.idairyj.2016.11.008

Ntuli, V., Njage, P.M.K., Buys, E.M., 2016. Characterization of Escherichia coli and other

Enterobacteriaceae in producer-distributor bulk milk. J. Dairy Sci. doi:10.3168/jds.2016-11403

Panchalingam, S., Antonio, M., Hossain, A., Mandomando, I., Ochieng, B., Oundo, J., Ramamurthy, T., Tamboura, B., Zaidi, A.K.M., Petri, W., Houpt, E., Murray, P., Prado, V., Vidal, R., Steele, D., Strockbine, N., Sansonetti, P., Glass, R.I., Robins-Browne, R.M., Tauschek, M., Svennerholm, A.M., Kotloff, K., Levine, M.M., Nataro, J.P., 2012. Diagnostic microbiologic methods in the GEMS-1 case/control study. Clin. Infect. Dis. 55, 294–302. doi:10.1093/cid/cis754

Platts-Mills, J.A., Babji, S., Bodhidatta, L., Gratz, J., Haque, R., Havt, A., McCormick, B.J.J., McGrath, M., Olortegui, M.P., Samie, A., Shakoor, S., Mondal, D., Lima, I.F.N., Hariraju, D., Rayamajhi, B.B., Qureshi, S., Kabir, F., Yori, P.P., Mufamadi, B., Amour, C., Carreon, J.D., Richard, S.A., Lang, D., Bessong, P., Mduma, E., Ahmed, T., Lima, A.A.A.M., Mason, C.J., Zaidi, A.K.M., Bhutta, Z.A., Kosek, M., Guerrant, R.L., Gottlieb, M., Miller, M., Kang, G., Houpt, E.R., Chavez, C.B., Trigoso, D.R., Flores, J.T., Vasquez, A.O., Pinedo, S.R., Acosta, A.M., Ahmed, I., Alam, D., Ali, A., Rasheed, M., Soofi, S., Turab, A., Yousafzai, A.K., Bose, A., Jennifer, M.S., John, S., Kaki, S., Koshy, B., Muliyil, J., Raghava, M.V., Ramachandran, A., Rose, A., Sharma, S.L., Thomas, R.J., Pan, W., Ambikapathi, R., Charu, V., Dabo, L., Doan, V., Graham, J., Hoest, C., Knobler, S., Mohale, A., Nayyar, G., Psaki, S., Rasmussen, Z., Seidman, J.C., Wang, V., Blank, R., Tountas, K.H., Swema, B.M., Yarrot, L., Nshama, R., Ahmed, A.M.S., Tofail, F., Hossain, I., Islam, M., Mahfuz, M., Chandyo, R.K., Shrestha, P.S., Shrestha, R., Ulak, M., Black, R., Caulfield, L., Checkley, W., Chen, P., Lee, G., Murray-Kolb, L.E., Schaefer, B., Pendergast, L., Abreu, C., Costa, H., Moura, A. Di, Filho, J.Q., Leite, ??lvaro, Lima, N., Maciel, B., Moraes, M., Mota, F., Ori??, R., Quetz, J., Soares, A., Patil, C.L., Mahopo, C., Mapula, A., Nesamvuni, C., Nyathi, E., Barrett,

L., Petri, W.A., Scharf, R., Shrestha, B., Shrestha, S.K., Strand, T., Svensen, E., 2015. Pathogen-specific burdens of community diarrhoea in developing countries: A multisite birth cohort study (MAL-ED). Lancet Glob. Heal. 3, 564–575. doi:10.1016/S2214-109X(15)00151-5

Rasko, D.A., Webster, D.R., Sahl, J.W., Bashir, A., Boisen, N., Scheutz, F., Paxinos, E.E., Sebra, R., Chin, C.-S., Iliopoulos, D., Klammer, A., Peluso, P., Lee, L., Kislyuk, A.O., Bullard, J., Kasarskis, A., Wang, S., Eid, J., Rank, D., Redman, J.C., Steyert, S.R., Frimodt-Møller, J., Struve, C., Petersen, A.M., Krogfelt, K.A., Nataro, J.P., Schadt, E.E., Waldor, M.K., 2011. Origins of the *E. coli* Strain Causing an Outbreak of Hemolytic–Uremic Syndrome in Germany. N. Engl. J. Med. 365, 709–717. doi:10.1056/NEJMoa1106920

Samie, A., Obi, C.L., Dillingham, R., Pinkerton, R.C., Guerrant, R.L., 2007. Enteroaggregative Escherichia coli in Venda, South Africa: Distribution of virulence-related genes by multiplex polymerase chain reaction in stool samples of human immunodeficiency virus (HIV)-positive and HIV-negative individuals and primary school child. Am. J. Trop. Med. Hyg. 77, 142–150.

Tau, N.P., Meidany, P., Smith, A.M., Sooka, A., Keddy, K.H., 2012. Escherichia coli o104 associated with human diarrhea, South Africa, 2004-2011. Emerg. Infect. Dis. 18, 1314–1317. doi:10.3201/eid1808.111616

World Bank, 2014. Economics of South African townships : special focus on Diepsloot. World Bank Studies. World Bank, Washington, D.C. doi:doi:10.1596/978-1-4648-0301-7