Journal of Water & Health

Journal of Water and Health Vol 20 No 7, 1137 doi: 10.2166/wh.2022.117

Potentially pathogenic *Escherichia coli* from household water in peri-urban Ibadan, Nigeria

Akeem Ganiyu Rabiu[a], Olutayo Israel Falodun[a,*], Obasola Ezekiel Fagade[a], Rotimi Ayodeji Dada[b] and Iruka N. Okeke[c]

[a] Department of Microbiology, University of Ibadan, Ibadan, Nigeria
[b] College of Health Sciences, Bowen University, Iwo, Osun State, Nigeria
[c] Department of Pharmaceutical Microbiology, University of Ibadan, Ibadan, Nigeria
*Corresponding author. E-mail: falod2013@gmail.com

AGR, 0000-0001-6232-7680; OIF, 0000-0002-6268-2532

ABSTRACT

Feco-orally transmitted infectious diseases are common in Nigeria where the potable water access is poor. In the south-western Nigerian Ibadan metropolis, supply of municipal water is meagre as residents depend on household wells and boreholes. The likelihood of fecal contamination of household water sources in Ibadan was examined longitudinally to quantify and understand its impact. Well and borehole water samples aseptically collected from 96 households in Ibadan were assessed for total heterotrophic counts (THCs), total coliform counts (TCCs) and total *Escherichia coli* counts (TECs) using a pour plate technique. *E. coli* were identified by *uidA* and whole-genome sequencing using Illumina technology, whereas virulence factors were predicted using VirulenceFinder. There was season-independent abundance of THC and TCC in the well and borehole with a significant recovery of *E. coli* in the wells during the wet season compared to the dry season ($P = 0.0001$). Virulence genes associated with pathogenic *E. coli* were identified in 13 (52%) strains with one *E. coli* each classified as extra-intestinal *E. coli*, avian pathogenic *E. coli* and enteroaggregative *E. coli*. High heterotrophic and coliform counts, with rainfall-driven *E. coli* contamination revealed that the water sources evaluated in this study are unfit for consumption.

Key words: *Escherichia coli*, geographic information system, household water, virulence genes

HIGHLIGHTS

- Spatio-temporal analyses of bacterial counts were used to investigate drinking water quality in Ibadan, Nigeria.
- Rainfall-driven fecal contamination revealed that well water sources were unfit for drinking.
- A novel *Escherichia coli* sequence type, ST13028, was found in a household well water.
- Potentially virulent extra-intestinal, avian pathogenic and enteroaggregative *E. coli* strains were recovered from household wells.

INTRODUCTION

Clean water and safe sanitation are prerequisites for good health but many people do not have these basic needs (World Health Organization/United Nation Children Education Fund 2006). In populations with a limited access to municipal water supply, use of alternative water sources such as wells and boreholes for domestic purposes is common (Aboh *et al.* 2015). Generally, these groundwater sources are among the most vital finite natural sources of fresh water on earth for drinking and presumed to be free from contamination (Avtar *et al.* 2019). However, due to watershed erosion, drainage from sewage, run-offs and non-enforcement of guidelines for groundwater exploration or well construction, bacterial contamination of groundwater sources does occur (Eduvie *et al.* 2003). These water sources place people at risk of water-borne diseases like diarrhea and typhoid (Kaplan *et al.* 2017). Consequently, around 6–8 million people die each year due to water-transmitted diseases (United Nation Water 2013). To estimate the scale of the problem and prevent diarrheal disease morbidity and mortality, it is necessary to investigate and monitor the microbiological quality of drinking water (Momtaz *et al.* 2013). Fecal contamination of water meant for human use constitutes a threat to public health since such water can be a vehicle to transmit diseases such as infectious diarrhea and typhoid.

Most *Escherichia coli* are commensals but some strains are pathogenic. Pathogenic strains are subdivided into intestinal pathogenic *E. coli* (InPEC), also known as diarrheagenic *E. coli* (DEC) and extra-intestinal pathogenic *E. coli* (ExPEC). The six well characterized InPEC pathotypes are Shiga toxin-producing *E. coli* (STEC), enteropathogenic *E. coli* (EPEC), enterotoxigenic *E. coli* (ETEC), enteroaggregative *E. coli* (EAEC), diffusely adherent *E. coli* (DAEC) and enteroinvasive *E. coli* (EIEC). These pathotypes were so categorized based on virulence genes and gastrointestinal pathogenesis mechanisms (Nataro & Kaper 1998; Kaper *et al.* 2004). ExPEC are classified based on the infection type into four groups: neonatal meningitis *E. coli* (NMEC), sepsis-associated *E. coli* (SEPEC), uropathogenic *E. coli* (UPEC) and avian pathogenic *E. coli* (APEC) (Riley 2014).

Urban Ibadan households largely rely on wells, springs and boreholes (Adesoji & Ogunjobi 2016) due to inability of the public water supply to meet population needs. Diarrheal disease and enteric fever incidence are not evenly distributed across the city and this has been attributed to differences in water safety and its microbiological quality, in particular (Popoola *et al.* 2019). Very little information is, however, available from the peri-urban areas. For instance, Akinyele, a peri-urban local government area (LGA) is not covered by the public water distribution network and for the most part, residents of the neighboring Lagelu LGA (LLGA) do not enjoy regular municipal water supply. Previous studies conducted in these areas on flooding, management of wells and drinking water assessment indicated poor water quality. However, these studies were performed through questionnaires or single sample analyses (Ince *et al.* 2010; Agbola *et al.* 2012) thus necessitating a longitudinal and laboratory-based investigation of the water sources.

With seasonal exacerbation of risk, we hypothesized that peri-urban household water sources constitute health risk for diarrhea and other enteric infections. Thus, we aimed to longitudinally identify enteric bacteria as a function of bacteriological quality of these sources. In addition, we also aimed to use the whole-genome sequencing (WGS) data to assess the *E. coli* virulence profiles since fewer studies have predicted pathotypes of *E. coli* recovered from household water sources.

METHODS

Study area

The study area is north-east of the Ibadan metropolis (Figure 1) in the south-western Nigeria. Peri-urban locations in what, until recently, were Akinyele and Lagelu LGAs were selected based on their limited or non-connectivity to Ibadan piped water (United Nation Economic and Social Council 2017). Due to urban renewal, the study sites were re-demarcated in the course of the study and are presently designated as follows – Akinyele LGA (ALGA), LLGA, Akinyele East local council development area, LCDA, (AELCDA), Akinyele South LCDA (ASLCDA), Lagelu North LCDA (LNLCDA) and Lagelu West LCDA (LWLCDA). Four of these six locations, namely ALGA, ASLCDA, AELCDA and LWLCDA are not included in the public water distribution network.

Study design and sampling

The total sample size (n) was estimated at 271 determined by the following equation

$X = Z\,(c/100)2r\,(100–r)$; $n = NX/((N-1)\,E2 + X)$; $E = \sqrt{\{(N-n)\,X/n\,(N-1)\}}$ using Raosoft (2004) software, with the following pre-estimated parameters, namely Error margin E, 5%; confidence level 90%; response distribution r, 50%; total estimated population, N, 505, 100 (National Population Commission 2015). Major assumptions are normal distribution and sample size (n) > 30. The study was planned longitudinally to collect household water samples over four seasons – dry and wet – for 2 years. However, three samples were collected each from 96 household water sources – well and borehole from January 2019 to March 2020. The intention to collect a fourth sample was thwarted by movement restrictions due to the COVID-19 pandemic. There is relative abundance of well over borehole water sources, thus, the well to borehole water sampling ratio was 2:1 and all samples were Global Positioning System (GPS)-located. The water samples were collected as previously described (Akinyemi *et al.* 2006), transported to the laboratory in a cool box and processed for microbiological analysis within 4 h of collection.

Spatio-temporal analyses of bacteria counts

Time-based analysis was done by aggregating data from weekly sample collection to the corresponding month. The mean monthly data were then analyzed through time series and seasonality using Excel. In the specifics, to inspect fluctuations over the sampling period, data were sequentially arranged per month over the study period of January 2019 to March 2020 for total heterotrophic, coliform and *E. coli* counts. As raw data can lead to misleading interpretation in geographical

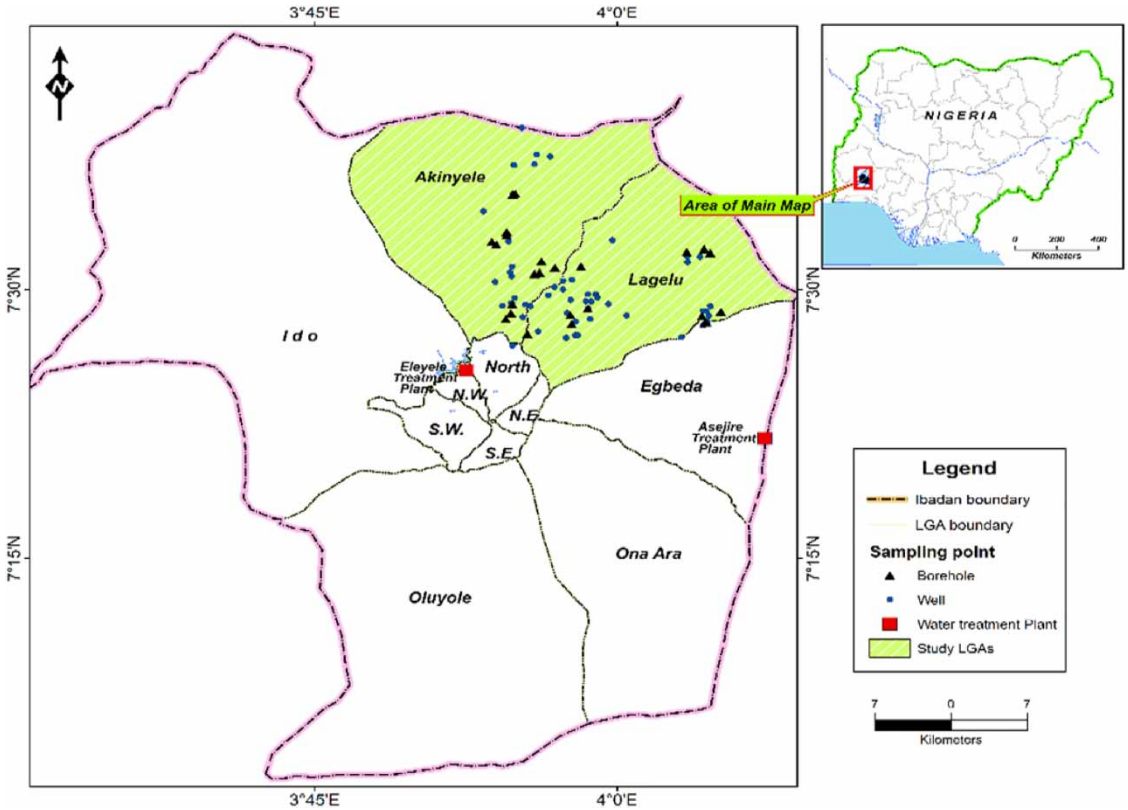

Figure 1 | A topographical map showing the study area (Nigeria).

areas with lesser populations (Owusu-Edusei & Owens 2009), the monthly mean data arranged based on corresponding month were subjected to exponential smoothing using Excel. The outcome was used to create data plus trend time series. Seasonality was determined by calculating the seasonal index which was computed by dividing the monthly average with the total sampling average for total heterotrophic, coliform and *E. coli* counts. The indices were then graphed by month (Figures 2–4). To achieve spatial analysis, sample level data (total heterotrophic/bacteria, coliform and *E. coli*) were added and averaged seasonally and allocated to each administrative area where the samples were collected. The average seasonal allocated data for total bacteria, coliform and *E. coli* were treated as point data and drawn into nine different maps using inverse distance interpolation method in the Arc Geographic Information System (GIS) platform (Figure 5).

Rainfall data

Estimates of Ibadan monthly rainfall from January to December (Ibadan Rainfall 2019) were sourced secondarily from the Nigeria Meteorological Agency (NiMet), which is publicly available (https://www.weather-nga.com/en/nigeria/ibadan-weather-january). These estimates were graphed to generate a rainfall time series for the city of Ibadan.

Direct plating and enrichment for enteric bacteria in water

The THC, TCC and TEC were obtained using recommended standard pour plate method (Wohlsen *et al.* 2006), with modifications. Briefly, THC and TCC were obtained on Nutrient agar (Oxoid) and MacConkey agar (Oxoid), respectively. In order to detect *E. coli* in the water samples, 1,000 μL of undiluted water was pour-plated with Eosine Methylene Blue (EMB) agar. The plates were observed for growth after incubation at 37 °C for 24 h. The plates were incubated overnight at 37 °C and observed for growth. Characteristic colonies were sub-cultured, after which the isolates were further identified by colonial, morphological and standard biochemical tests. Green metallic sheen on EMB agar was used for *E. coli* pre-estimation. The *E. coli* isolates were stored in Luria Bertani-Glycerol (LB Glycerol) broth and frozen at −80 °C until further analysis.

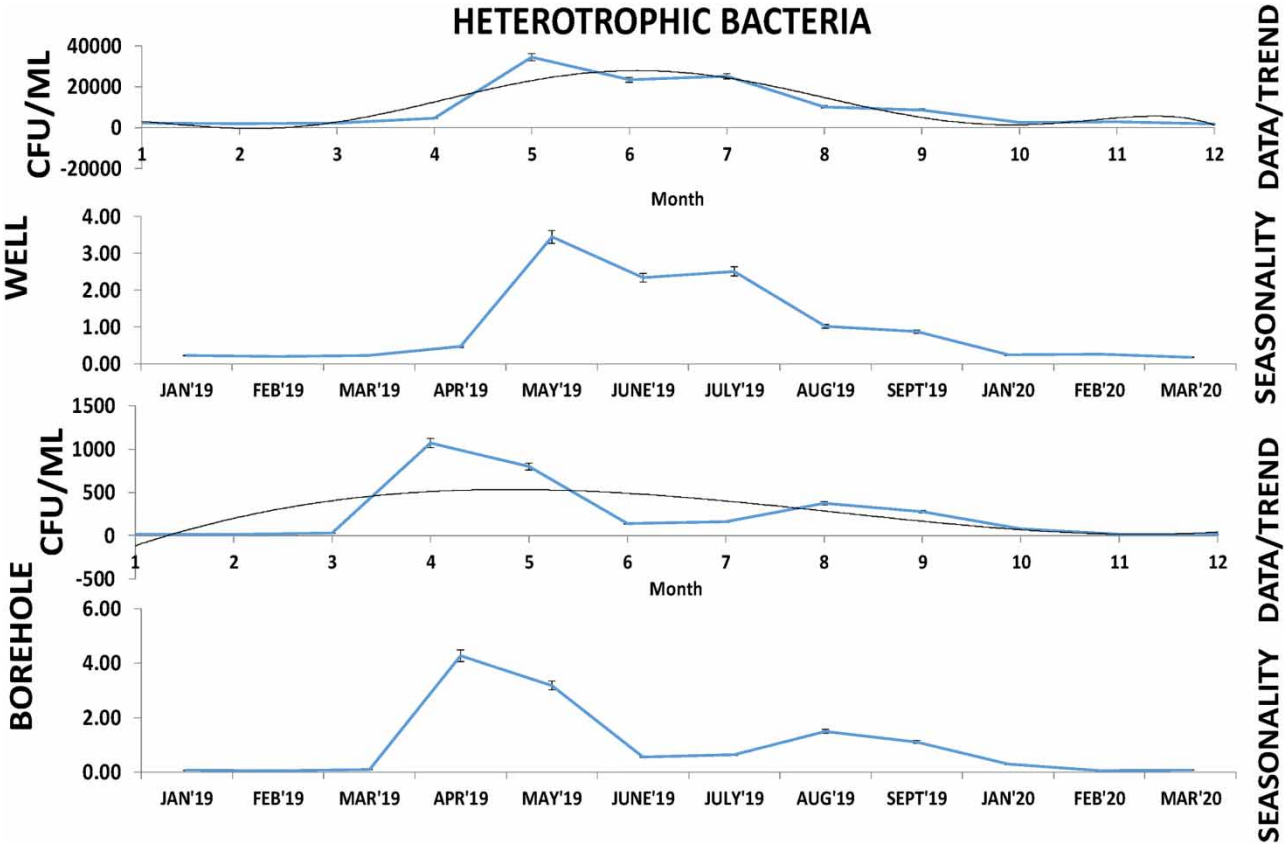

Figure 2 | Time-series showing well water THBC raw data/trend in the first graph, well THBC seasonality in the second graph, borehole water THBC raw data/trend in the third graph and borehole THBC seasonality in the fourth graph. The figure represents data collected during 2019 (dry and wet) and 2020 (dry) seasons from Akinyele and Lagelu regions of Ibadan, Nigeria.

Extraction of genomic DNA, gDNA

The *E. coli* isolates were cultured in Luria Bertani broth on a shaker for 24 h at 37 °C. The genomic DNA (gDNA) was extracted using Wizard® Genomic DNA Purification Kits (Promega) following the manufacturer's procedure.

E. coli confirmation

PCR was performed on extracted gDNA to detect sequences specific for *E. coli* housekeeping *uidA* (*β*-D glucuronidase) gene (Divyashree *et al.* 2015) with the forward UAL-754 5′AAAACGGCAAGAAAAAGCAG 3′ and reverse primers UAR-900 5′CGCGTGGTTACAGTCTTGCG 3′ at 146 bp (base pairs). Reactions were performed in 20 μL total reaction volumes containing 10 μL One Taq Quick-Load 2 × Master Mix with Standard Buffer (New England, BioLabs). Amplification was conducted in a programmable thermal cycler (Applied Biosystems Veriti 96 Well Thermal Cycler). The reaction was carried out using 1.0 μL of DNA template, 0.1 μL of each forward and reverse primers, and 8.8 μL nuclease-free water. The PCR program was optimized to consist of an initial denaturation at 94 °C for 5 min followed by 30 cycles at 94 °C for 30 s, annealing at 60 °C for 30 s and extension at 72 °C for 30 s with final extension at 72 °C for 10 min. *E. coli* ATCC 25922 was used as positive control while DNase-free molecular grade water was used as a negative control.

All the PCR products were electrophoresed on 1.5% agarose (w/v) gel at 150 V, 277 mA for 50 min. Amplicon size was estimated with a 100 bp molecular weight marker (Invitrogen) and the products were visualized using an ultraviolet trans-illuminator (UVP GelMax Imager).

Whole-genome sequencing

To determine the pathotypes of the 25 *E. coli*, and their inter-relationships, if any, WGS was carried out using Illumina technology. The gDNA standardized to 0.2 ng/μL was submitted to paired-end WGS using MiSeq sequencer (Illumina, San Diego,

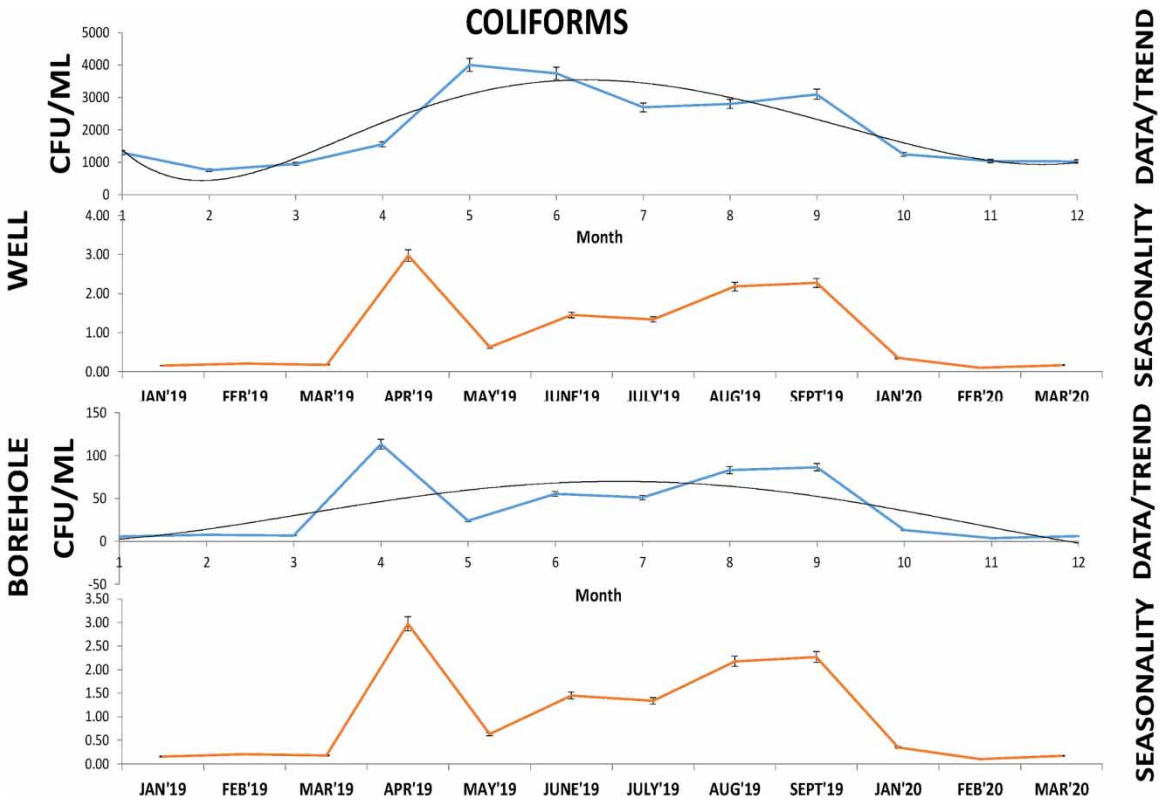

Figure 3 | Time-series showing well water TCC raw data/trend in the first graph, well TCC seasonality in the second graph, borehole water TCC raw data/trend in the third graph and borehole TCC seasonality in the fourth graph. The figure represents data collected during 2019 (dry and wet) and 2020 (dry) seasons from Akinyele and Lagelu regions of Ibadan, Nigeria.

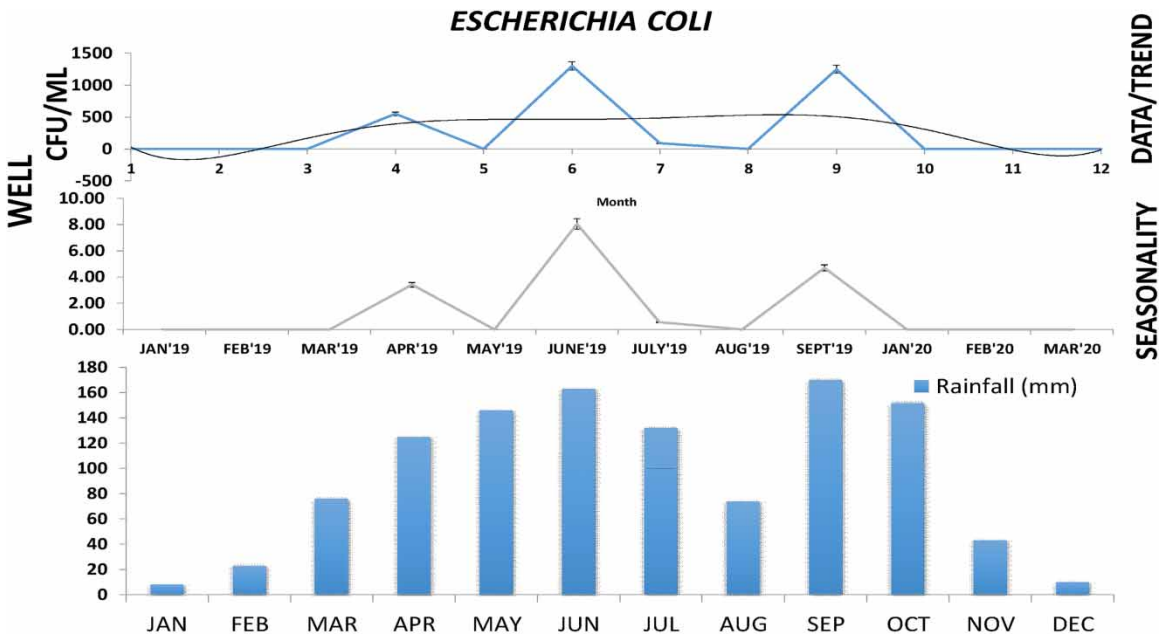

Figure 4 | Time-series showing well water TEC raw data/trend in the first graph, well TEC seasonality in the second graph and rainfall data from January to December, 2019 are shown in the third graph. *Escherichia coli* was not found in the borehole water sources. The figure represents data collected during 2019 (dry and wet) and 2020 (dry) seasons from Akinyele and Lagelu regions of Ibadan, Nigeria.

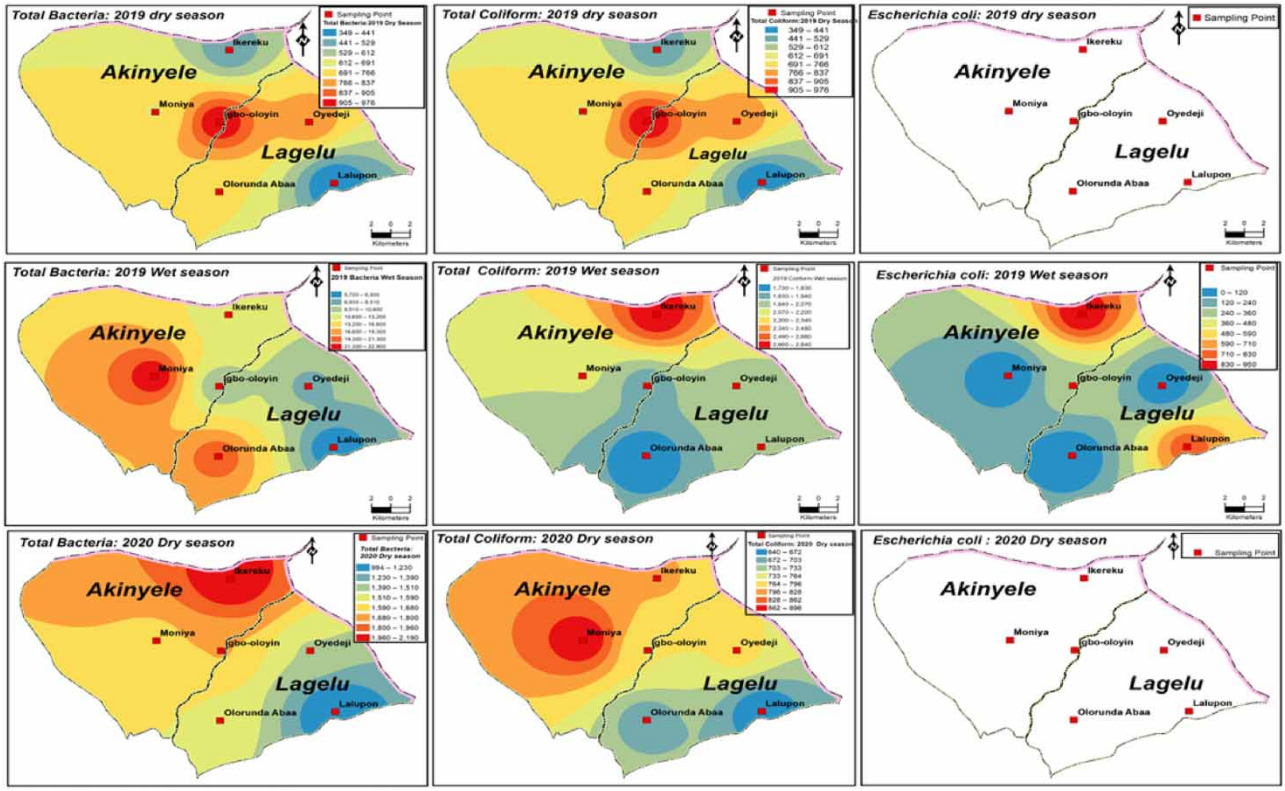

Figure 5 | Total heterotrophic bacteria, coliform and *Escherichia coli*-positives at the sampling level using seasonal mean data by administrative area in Akinyele and Lagelu regions of Ibadan for the years 2019 (wet and dry seasons) and 2020 (dry season). The 2019 dry season THBC is shown in the first column + first row; the 2019 wet season THBC is shown in the first column + second row, whereas the 2020 dry season THBC is shown in the first column/third row. The 2019 dry season TCC is shown in the second column + first row, the 2019 wet season TCC is shown in the second column + second row, whereas the 2020 dry season TCC is shown in the second column + third row. Finally, the third column + first row showed the TEC for the 2019 dry season, the third column + second row showed the 2019 wet season TEC, whereas the third column + third row showed the 2020 dry season TEC.

CA, USA). Genomic libraries were prepared using the Nextera XT kit (Illumina, CA, DNA). The gDNA was fragmented and tagged with sequencing adapters in a single step using Nextera transposome (Nextera XT DNA Library Preparation Kit, Illumina, San Diego, USA). The fragmented and tagged gDNA was amplified (PCR 12-cycle amplification) and cleaned up with AMPure beads. Nextera libraries were measured using Qubit and the size profile was analyzed on 2100 Bio-analyzer using high-sensitivity DNA assay kit (Agilent Technologies, Waldbronn, Germany). Libraries selected for sequencing were normalized to 1 nM and pooled. The 1 nM pooled library was denatured and diluted prior to loading on a MiSeq paired-end 2 × 150 bp (MiSeq reagent kit V2 (300 cycles)) or 2 × 300 bp (MiSeq reagent kit V3 (600 cycles)) sequence run.

Genome assembly, species identification and phylogenetic tree

Following from WGS, raw sequences were submitted to fastqc and multiqc to assess and aggregate phred scores and cytosine to Guanine (CG) content. The contigs were then annotated using Prokka v1.12 with contigs <200 bp and coverage <10-fold excluded from the analyses. The assemblies were checked for quality strictures while the assembled reads were aligned to *E. coli*-042 reference genome to create an alignment file using Bcftools, Samtools, Gatk, Bwa, Smalt and Picard. The single nucleotide polymorphism (SNP) was performed on the alignment file using SNP sites to produce the SNP file. The SNP file was used to construct phylogenetic tree using IQ tree tool and the tree was viewed with interactive tree of life (iTOL). The species were determined using Pathogenwatch resource available at https://pathogen.watch/upload/fasta. Genomes downloaded from the European Nucleotide Archive (ENA) database and incorporated into phylogenetic tree are SRR6362403, a soil environment *E. coli* – Minnesota, USA (North Dakota State University 2015); ERR987377, an enteroaggregative *E. coli* clinical isolate in Nigeria (Dougan *et al.* 2018); SRR11654221, a food animal *E. coli* – USA (United States

Department of Agriculture, Food Safety and Inspection Service 2020) and FN554766, an EAEC_042 – Human *E. coli* isolate (Chaudhuri *et al.* 2010).

Screening for virulence-associated genes and *E. coli* phylotyping

Virulence-associated genes (VAGs) were identified by the center for genomic epidemiology resources tool VirulenceFinder https://cge.cbs.dtu.dk/services/VirulenceFinder/ that includes 38 ExPEC-associated virulence genes (1,890 alleles) recently added to the existing *E. coli* VirulenceFinder database (Tetzschner *et al.* 2020). Phylogroups were identified through the ClermonTyper with fasta files uploaded to http://clermontyping.iame-research.center/. Multilocus sequence typing (MLST) was performed by the Achtman MLST scheme that uses seven housekeeping genes (*adk*, *fumC*, *gyrB*, *icd*, *mdh*, *purA* and *recA*) through the website https://cge.cbs.dtu.dk/services/MLST/.

Assignment of number to novel sequence type

In order to assign ST number to novel STs, the assembled sequences (fasta) of the isolates were first uploaded to the Centre for Genomic Epidemiology platform at MLST – Typing Results (dtu.dk) using MLST-2.0 Server where allele numbers were given. Then, the raw sequences (fastq) were uploaded to Enterobase – Search ecoli (warwick.ac.uk) and allele numbers inputed to yield ST number.

Statistical analysis

Data were subjected to Pearson correlation and paired sample *t*-test (IBM SPSS 20 version) to show relationship among the microbiological parameters at levels of 5 and 1%. Biological characteristics of the source water over the seasons were compared using Fischer's exact test.

RESULTS

Household water bacteria quality

Of the 96 household water sources sampled each season from Akinyele and Lagelu communities, 66 samples were collected from wells while 30 were collected from boreholes. The number of wells with THC ≥500 CFU/mL and TCC ≥10 CFU/mL in wet (THC = 63, TCC = 60) and dry seasons (THC = 57, TCC = 59) was comparable (Supplementary Material, file S1). Five and 12 boreholes, respectively, had THC ≥500 CFU/mL and TCC ≥10 CFU/mL in wet season. No borehole had THC ≥500 CFU/mL but 10 boreholes had TCC ≥10 CFU/mL in dry seasons (Table 1). Thirteen (19.70%) of the 66 wells were contaminated with *E. coli* during the wet season. However, borehole samples did not have *E. coli* contamination in either season (Table 1). The water sources showed season-independent heterotrophic contamination, THC; $P = 0.127$ (well) and $P = 0.052$ (borehole), coliform contamination, TCC ($P = 0.492$ (well) and $P = 0.789$ (borehole)). The *E. coli* contamination, TEC, for well water sources was found to be significant ($P = 0.0001$) across the seasons (Table 1).

Spatio-temporal analyses of enumerated bacteria

The number of samples collected and submitted to total bacteria, coliforms and *E. coli* testing per months were higher in dry seasons than the wet season (Supplementary Material, files 2S1, 2S2 and 2S3). This is because dry samples were collected in 3

Table 1 | Biological characteristics of household water over seasonal changes in peri-urban Ibadan using Fischer's exact test

	Well			Borehole		
Estimated parameters	***Dry***	**Wet**	***P*-values**	***Dry***	**Wet**	***P*-values**
THC <500 CFU/mL	9	3	0.127	30	25	0.052
THC ≥500 CFU/mL	57	63		0	5	
TCC <10 CFU/mL	7	3	0.492	20	18	0.789
TCC ≥10 CFU/mL	59	63		10	12	
TEC <1 CFU/100 mL	66	53	0.0001	0	0	
TEC ≥1 CFU/100 mL	0	13		0	0	

Dry is the average positive water samples for the two dry seasons. No *E. coli* was detected in the borehole samples.
THC, total heterotrophic count; TCC, total coliform count; TEC, total *E. coli* count.

months from January through March while wet samples were collected from April to September (6 months). In order to give a clearer picture of how seasonal variation especially rainfall drives contamination of water sources, the graphed annual rainfall data from January to December were aligned with the total heterotrophic, coliform and *E. coli* time series and seasonality for wells and boreholes. The time series and seasonality plot for total bacteria and coliforms showed two peaks in April/May and around July to September. The result showed the fluid impact of rainfall on groundwater contamination. The higher total heterotrophic and coliform counts correspond to the wet season as shown in June and September rainfall peaks. However, *E. coli* time series and seasonality plot specifically showed peaks in April, June and September. The use of the mean data as point data for each location (Supplementary Material, files 3S1 and 3S2) to create maps (total bacteria, coliforms and *E. coli*) showed fluctuations depicted in a south to north and east to west gradient in total bacteria and coliforms. The area with high *E. coli* contamination with 700–950 CFU/mL are localized to northern part (Ikereku) and south-eastern locales (Lalupon) of the study area. For these two locations, the total coliform maps where coliforms were greater than or equal to 2,000 CFU/mL is consistent with *E. coli* contamination (Figure 5).

Phylogenetic relationship and MLST

The 25 *E. coli* obtained in this study belonged to three phylogroups. Eleven (44%) belonged to phylogroup A, 13 (52%) to group B1, while 1 (4%) belonged to group D. The isolates of this study, namely 5EC, 108EC, 172EC and the *E. coli* genome with the accession number ERR987377 obtained from European Nucleotide Archive, ENA belonged to ST10. Other ENA strains – FN554766, SRR6362403, SRR11654221 belonged to ST414, ST9005 and ST11, respectively. The 13EC, 14EC, 15EC and 16EC belonged to ST2541, whereas 171EC, 172EC1, 173EC and 175EC belonged to ST2008. ST156 included 69EC, 70EC and 70EC1 strains, ST48 included 166EC and 184EC strains but 167EC and 167EC1 belonged to a novel ST13028. Other strains belonged to the following STs-58, 69, 155, 196, 4981, 11400 and 9987*.

Concurrent analysis of the STs and phylogroup showed that all the ST10, ST2541, ST48, ST11400 and ST4981 isolates belonged to the phylogroup A as against ST2008, ST196, ST155, ST58, ST156 and ST9005 that belonged to the phylogroup B1. The only ST69 isolate of this study belonged to phylogroup D as did enteroaggregative *E. coli* FN554766 (ST414) retrieved from ENA. The two strains, ST/phylogroup, ST69/group D and ST414/group D shared certain identical set of virulence genes. As indicated in the phylogrouping, the strains spanned a broad range of known *E. coli* phylogeny. Isolates that appeared to be clonal, based on having no or very few SNP differences between them and identical virulence gene profiles were isolated from nearby water sources in the wet seasons (Figure 6, Supplementary Material, files 4S1 and 4S2).

Analysis of the VAGs

Analysis of the VAGs showed that majority of the strains did not carry virulence genes usually possessed by diarrheagenic pathotypes. The total number of different VAGs harbored by the strains were 33. The *E. coli* strains possessed an average of six virulence genes (range 3–16) per isolate. ExPEC classification requires possession of the following virulence genes – *papA/papC*, *sfa/foc*, *Afa/dra*, *kpsM II*, *ompT* and *iutA*, whereas EAEC is inclusive of *eilA, chuA, fyuA* and *ompT* genes. Among other VAGs, APEC subpathotype possesses *iutA, iss*, *cvaC*, *IroN* and *ompT* genes. Based on the virulence gene repertoire, three (12%) were potential ExPEC (Johnson *et al.* 2003), EAEC (Chaudhuri *et al.* 2010) or APEC (Collingwood *et al.* 2014), the rest were classified as likely commensals. With the exception of three isolates (191EC, 166EC, 184EC), all the strains possessed an average of 1–3 adhesin genes (*afaD*, *afaA*, *iha*, *hra* and *lpfA*). The strain 172EC possessed *AfaA/D* typical of diffusely adherent *E. coli*. Among the adhesin genes were *iha,* EHEC Novel non-hemagglutinin adhesin from O157:H7 (Tarr *et al.* 2000) and *hra2*, heat-resistant agglutinin gene that encodes integral outer membrane hemagglutinin (Mancini *et al.* 2011) possessed by 172EC strain that met the ExPEC classification. The long polar fimbrial structural subunit gene *lpfA* was contained in 18 (72%) of the *E. coli* genomes.

Two isolates carried toxin-related genes such as *Shigella* enterotoxin B gene (*senB*), *astA*, hemolysin, colicin and/or microcin (*hlyF–*, *cia*, *cba*, *cib*, *cvaC* and *cma*). The isolate 191EC carried *hlyF-*, *cia*, *cba*, *cib*, *cvaC* and *cma*, whereas 172EC carried *senB*, *astA* and *hlyF* toxin-associated virulence genes, though *senB* and *astA* are also seen in DEC. An average of two genes responsible for bacteria iron metabolism such as iron acquisition, yersiniabactin biosynthesis and aerobactins genes; *iroN*, *irp2*, *fyuA*, *chuA*, *iutA*, *iucC–* and *traT–* were found in nine *E. coli* strains. The *irp2* and *fyuA* genes that encode yersiniabactin biosynthesis and receptor protein from *Yersinia* Pathogenic Island (PAI) were possessed by seven isolates as did surface exclusion, *traT* genes. The capsular genes *kpsE* and *kpsMIII* were found in two strains; one (154EC) carried *kpsE* only while the other (183EC) carried both.

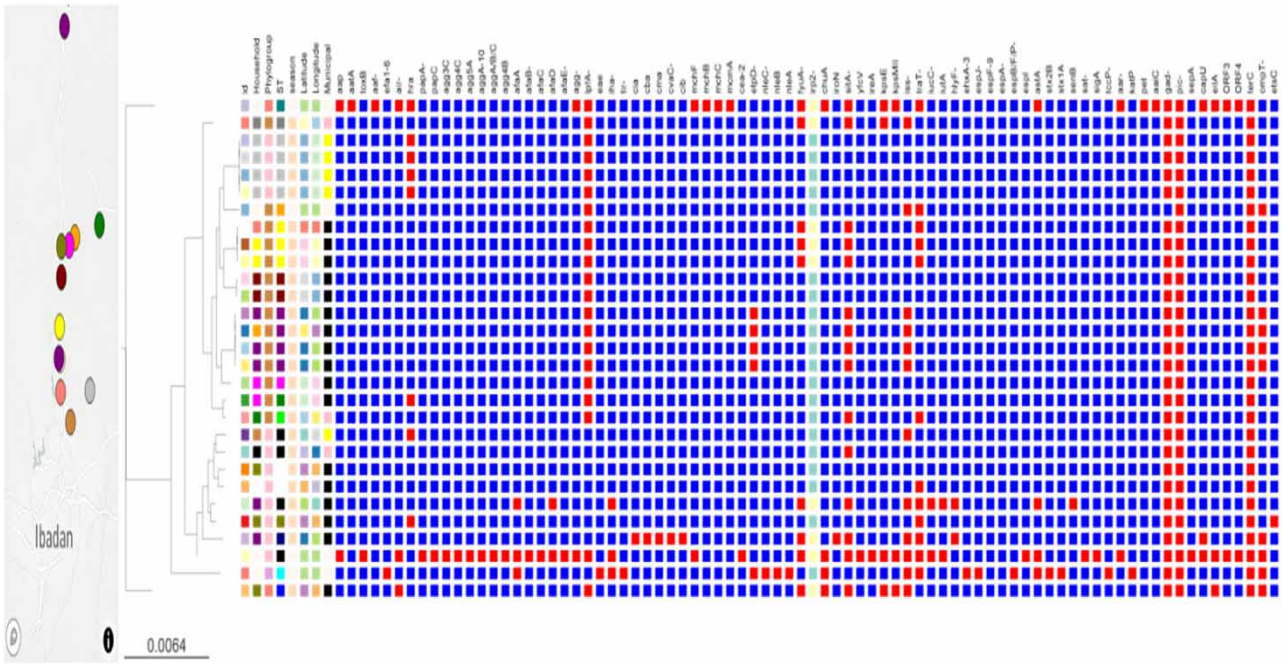

Figure 6 | SNP-based cladogram of 25 *E. coli* recovered from well water based on reference genome *E. coli*_042. Each household was given a unique color and the isolates assigned to the household where they were isolated to illustrate incidence of clonal and non-clonal strains across and within households. On the left side of the cladogram, the households 1–13 were, respectively, colored *black, silver, gray, maroon, green, salmon, yellow, purple, olive, orange, fuchsia, white, peru.* Please refer to the online version of this paper to see this figure in color: http://dx.doi.org/10.2166/wh.2022.117.

Additional ExPEC-associated virulence factors include glutamate decarboxylase *gad*, an acid survival factor and tellurite ion resistance *terC* genes that were possessed by all the strains. Thirteen of these isolates were positive to iron acquisition or transport system *sitA*, which encodes periplasmic-binding protein, whereas four of the latter possessed plasmid-borne type II secretion system genes, *etpD* seen in ETEC. Enteroaggregative immunoglobulin repeat protein *air* and pAA plasmid-encoded *eilA*, a *HilA*-like regulator EAEC was carried by 183EC classified as EAEC. Others were outer membrane protein *ompT* and increased serum survival *iss* genes respectively found in six and nine strains. Efflux transporter gene *etsC* and EAEC *capU* were possessed by one isolate respectively (Figure 6).

VAG profiles were used to classify the *E. coli* strains into pathotypes. One *E. coli* strain each was classified as ExPEC (172EC), EAEC (1823EC) and APEC (191EC). *E. coli* strains not classifiable into any pathotype were regarded as commensals. The Microreact visualization of the virulence genes and metadata are available at https://microreact.org/project/hGKMbAnF2REcU7VZv3SUgc-whole-genome-sequence-analysis-of-e-coli-obtained-from-household-well-water-sources-in-peri-urban-ibadan-nigeria.

Incidence of *E. coli* across and within households

SNP was used to investigate the point-source contamination of the well water sources which also infers the diversity and clonality of the *E. coli* strains in the 13 household water sources where the isolates were recovered. Across the households in the same municipal area (AELCDA); the strains 171EC, 173EC and 175EC were clonal (SNP = 0) even though 171EC and 173EC strains were found in household 8 as against 175EC strain found in household 10. Within the households, four (4) strains (13EC, 14EC, 15EC and 16EC) in the same sampling found in household 2 in the ASLCDA municipal area slightly diverged from one another by two to four nucleotides.

In household 7, the clonality of isolates (69EC and 70EC) recovered (SNP = 0) were similar to what occurred in household 8 where 171EC and 173EC were isolated. The third strain (172EC) also obtained in household 8 had wide genome difference (SNP = 26,816) from 171EC and (SNP = 26,757) from 173EC. In household 9, the three strains (182EC, 183EC and 184EC) obtained from the household well water sources were non-clonal as they had wide nucleotide difference that ranged from 10,487 to 47,637. Strains that were non-clonal ($n = 2$) found in household 11 also had extensive genome difference

(SNP = 13,679). The strains recovered in households 7, 8, 9 and 11 were, however, isolated within the same municipal area (AELCDA) (Figure 6, Supplementary Material files 4S1 and 4S2).

DISCUSSION

The increasing stress on available water resources worldwide and potential for disease spread require sentinel surveillance of water systems (Invik *et al.* 2017). It is also important to understand temporal variations in water contamination to effectively predict possible outbreaks, delineate high-risk periods (Hay *et al.* 2013; Galway *et al.* 2015) and raise evidence for clean water provision policy. Natural water sources are predisposed to chemical and bacteriological contaminants principally from animal and human waste (Nigerian Standard for Drinking Water Quality 2015) such as pit latrines, domestic waste and leakages from landfills that permeate peri-urban landscapes.

In Nigeria, water intended for domestic use is expected to be free from coliforms above 10 CFU/mL as such water is considered fecally contaminated (NSDWQ 2015). This study found that peri-urban well and borehole from Akinyele and Lagelu households were highly contaminated with bacteria and coliforms. The findings of this study are similar to reports from previous studies in Ibadan (Adesoji & Ogunjobi 2016), as well as other Nigerian cities such as Lagos (Akinyemi *et al.* 2006) and Zaria (Aboh *et al.* 2015). Furthermore, water intended for drinking should contain no detectable *E. coli* per 100 mL sample (WHO 2011; NSDWQ 2015). Standard pour plate and Petri film methods are known to give better and more consistent results than membrane filtration, defined substrate technology, Colilert-18, Quanti-Tray/2000 system and the most probable number (MPN) techniques which produced low recovery or significant variability between replicates in side-by-side evaluations (Wohlsen *et al.* 2006). In this study, we therefore used standard pour plate method and this yielded higher total *E. coli* counts of 4,250 CFU/100 mL than 16 CFU/100 mL reported elsewhere in Nigeria (Aboh *et al.* 2015).

Our observation that 20 (66.7%) of sampled boreholes had total coliform counts below 10 CFU/mL and no detectable *E. coli* (0 CFU/100 mL) demonstrates that this source was superior to wells, for which only seven (10.6%) were ever found to be potable. This is similar to what was reported in Limpopo, South Africa where Molekoa *et al.* (2019) found more than 80% of borehole water samples fit for drinking and domestic purposes. Boreholes are expensive to install and therefore beyond the reach of many households and communities in Ibadan. However, the poor quality of well water recorded and the high risk of contracting typhoid in Ibadan (Popoola *et al.* 2019) suggests that these safer water sources may be more cost-effective overall. As the boreholes did not, however, yield unanimously potable water, even boreholes cannot be presumed to be good sources of water, until assessments have been conducted.

Maps and time series provide more granular information on how temporality and geography affect groundwater contamination. For spatial analysis, we used GPS coordinates which permitted good resolution and visualization of the study area showing high-risk locales (Invik *et al.* 2017). *E. coli* contamination of wells in April, June and September corresponds with onset and abundance of rainfall as shown in the June and September rainfall peak (Figures 2–4) as similar observation was not recorded in the preceding and succeeding dry seasons. *E. coli*-positive wells have been previously reported in Nigeria (Aboh *et al.* 2015; Adesoji & Ogunjobi 2016), our study reveals that *E. coli* contamination of wells in Lagelu (Lalupon) and Akinyele (Ikereku) regions of Ibadan are driven by rainfall (Figure 4). Risk could also be worsened by factors that we did not track such as well shallowness, open defecation and agricultural practices that include use of animal manure in farmlands. For future surveillance in the general area, these maps (Figure 5) can be used to anticipate hot spots of microbial contamination in groundwater (Invik *et al.* 2017).

E. coli presence in water is a marker of fecal contamination, highlighting risk that harder-to-detect pathogens may be present in the water. However, some *E. coli* are pathogenic and therefore pose a risk in themselves. We examined the lineages and virulence factors of *E. coli* isolates from this study to determine whether they are potential pathogens. We used WGS because it offers the best resolution for comparison of pathogen genomes and exhaustively identifying virulence factors. Based on phylogrouping, *E. coli* is classified into eight groups – A, B1, B2, C, D, E, F or cryptic clade I (Clermont *et al.* 2000). The *E. coli* recovered in this study belong to phylogroup A, B1 and D classified as commensals (A and B1), pathogenic *E. coli* (A, B1 or D) and ExPEC (group D) (Clermont *et al.* 2000). Phylogroup B1 is commonly adapted to water environments and it predominated in the *E. coli* population in this study, i.e. B1 (52%), A (44%) and D (4%). This is similar to the observation of Nowicki *et al.* (2021) that used WGS to examine the association of *E. coli* with recent fecal contamination in Kenyan rural drinking water.

The MLST provides finer subtyping for *E. coli* than broad phylotyping as it gives clearer understanding of the spread and diversity of strains. Three ST10 isolates were recovered from different zones of the sampling region. The ST10 is a large and diverse clonal complex that includes human isolates and is the most common commensal detected in Africa, Asia, Europe and South-America (Rashid *et al.* 2015; Matamoros *et al.* 2017). The current study detected ST10 isolates in Moniya (172EC), Lalupon (108EC) and Arulogun (70EC) communities separated by not less than 20 km apart. This observation is in agreement with the dominance of ST10 lineages within the same setting, though among enteroaggregative *E. coli* especially from healthy individuals (Okeke *et al.* 2010). The predominant STs identified in this study were ST2541 and ST2008 as both constituted 32% of the isolates, including two near-identical ST2541 strains (1 SNP distance) that were isolated from the same well water point in the same sampling even though this ST is yet to be reported from water sources. One of the four ST2008 isolates, ID/ST, 172EC1/ST2008 has a wide, 2452, SNP distance from others. However, the other three ST2008 were clonal (SNP distance = 0) even though they were recovered from two different households (above 5 km apart) within the same municipal area (AELCDA) in the same season. It is significant to highlight that the current study reports a novel ST13028 yet to be reported in the literature.

One EAEC isolate was detected in the study and belonged to ST69, which like the *E. coli*_042 ST414 is in Phylogroup D (Okeke *et al.* 2010). The EAEC-ST69 isolate carried a similar virulence gene repertoire with Peru EAEC/*E. coli*_042 strain (Chaudhuri *et al.* 2010) that causes diarrhea. The ST48, also detected, has been reported from drinking water in France (Madec *et al.* 2016) as ST69, ST156 and ST10 detected in this study have been found in river water sources (Colomer-Lluch *et al.* 2013; Rashid *et al.* 2015). The ST10, ST155 and ST156 *E. coli* lineages are known to be linked to human *E. coli* isolate (Matamoros *et al.* 2017; Zahra *et al.* 2018) and this indicates cross-niche detection. Similarly, the only ExPEC isolate detected in this study that belonged to phylogroup A are known to cause infections in anatomical sites and sepsis (Nataro & Kaper 1998).

Lugli *et al.* (2022) recently reported tap water in Italy as a reservoir of bacteria that can colonize human intestine, shape gut microbiota and enable diverse metabolic processes attributed to human health. In the low and middle income countries without municipal water services, the microbiological quality of informal drinking water sources has been less thoroughly investigated (Kaplan *et al.* 2017; Nowicki *et al.* 2021). However, the few studies performed have yielded bacteria with pathogenic potential, including *Salmonella enterica* serovar Typhi, *S.* Paratyphi A and *E. coli*, although the virulence of *E. coli* isolates is rarely investigated. By evidencing high total and coliform counts and recovery of potentially pathogenic *E. coli*, this study adds to the evidence pointing to unsafe household water in Ibadan, and its potential for exacerbating feco-oral disease transmission.

CONCLUSIONS

Peri-urban Ibadan populations that are dependent on wells and boreholes with limited connection to municipal water works in Ibadan, investigated in this study do not have access to safe water as their water sources are largely contaminated with coliforms beyond the acceptable limits. Seasonal variations worsen the quality of the groundwater sources in the wet season. Overall, borehole water sources appeared to be safer, but not uniformly so. There is also a need to further assess these sources for other quality parameters before declaring the few sources without coliforms and *E. coli* as safe. *E. coli* recovery from these water sources include potential pathogens such as ExPEC, EAEC and APEC that pose a direct threat of disease spread.

ACKNOWLEDGEMENTS

This work is supported by the African Research Leader's Award, awarded to O.I.N. funded by the UK Medical Research Council (MRC) and the UK Department for International Development under the MRC/DFID Concordant agreement and is also part of the EDCTP2 program supported by the European Union. I.N.O. is a Calestous Juma Science Leadership Fellow supported by the Bill and Melinda Gates Foundation. The funders had no role in the content, crafting or submission of this paper.

AUTHORS' CONTRIBUTIONS

O.I.F., O.E.F., A.G.R. and I.N.O. conceived and designed the experiments. A.G.R. performed the experiments. O.I.F., I.N.O. and A.G.R. did literature search, and contributed reagents and materials. O.I.F., I.N.O., R.A.D. and A.G.R. analyzed the data and contributed analysis tools. A.G.R. wrote the first draft. All authors approved the final manuscript.

COMPETING INTERESTS

The authors have declared that no competing interests exist.

DATA AVAILABILITY STATEMENT

All relevant data are included in the paper or its Supplementary Information.

CONFLICT OF INTEREST

The authors declare there is no conflict.

REFERENCES

Aboh, E. A., Giwa, F. J. & Giwa, A. 2015 Microbiological assessment of well waters in Samaru, Zaria, Kaduna, State, Nigeria. *Ann. Afr. Med.* **14** (1). doi:10.4103/1596-3519.148732.

Adesoji, A. T. & Ogunjobi, A. A. 2016 Detection of extended spectrum beta-lactamases resistance genes among bacteria isolated from selected drinking water distribution channels in southwestern Nigeria. *BioMedRes. Int.* http://dx.doi.org/10.1155/2016/7149295.

Agbola, B. S., Owolabi-Ajayi, O., Olalekan, J., Taiwo, O. J. & Wahab, B. W. 2012 The August 2011 flood in Ibadan, Nigeria: anthropogenic causes and consequences. *Int. J. Disaster Risk Sci.* **3** (4), 207–217. doi:10.1007/s13753-012-0021-3.

Akinyemi, K. O., Oyefolu, A. O. B., Salu, O. B., Adewale, O. A. & Fasure, A. K. 2006 Bacterial pathogens associated with tap and well waters in Lagos, Nigeria. *East Cent. Afr. J. Surg.* **11** (1), 110–117.

Avtar, R., Tripathi, S., Aggarwal, A. K. & Kumar, P. 2019 Population-urbanization-energy Nexus: a review. *Resources* **8**, 136.

Chaudhuri, R. R., Sebaihia, M., Hobman, J. L., Webber, M. A., Leyton, D. L., Goldberg, M. D., Cunningham, A. F., Scott-Tucker, A., Ferguson, P. R., Thomas, C. M., Frankel, G., Tang, C. M., Dudley, E. G., Roberts, I. S., Rasko, D. A., Pallen, M. J., Parkhill, J., Nataro, J. P., Thomson, N. R. & Henderson, I. R. 2010 Complete genome sequence and comparative metabolic profiling of the prototypical enteroaggregative *Escherichia coli* strain 042. *PLoS ONE* **5**, E8801–E8801.

Clermont, O., Bonacorsi, S. & Bingen, E. 2000 Rapid and simple determination of the *Escherichia coli* phylogenetic group. *Appl. Environ. Microbiol.* **66**, 4555–4558.

Collingwood, C., Kemmett, K., Williams, N. & Wigley, P. 2014 Is the concept of avian pathogenic *Escherichia coli* as a single pathotype fundamentally flawed? *Front. Vet. Sci.* **1** (5). doi: 10.3389/fvets.2014.00005.

Colomer-Lluch, M., Mora, A., López, C., Mamani, R., Dahbi, G., Marzoa, J., Herrera, A., Viso, S., Blanco, J. E., Blanco, M., Alonso, M. P., Jofre, J., Muniesa, M. & Blanco, J. 2013 Detection of quinolone-resistant *Escherichia coli* isolates belonging to clonal groups O25b:H4-B2-ST131 and O25b:H4-D-ST69 in.raw sewage and river water in Barcelona, Spain. *J. Antimicrob. Chemother.* **68** (4), 758–765.

Divyashree, M., Shama, P. K., Veena, S. A. & Karunasagar, I. 2015 Antibiotic resistance in extended-spectrum beta-lactamase-producing *Escherichia coli* isolated from effluents of tertiary care hospitals. *Int. J. Health Sci. Res.* **5** (12), 27–33.

Dougan, G., Thomson, N. R. & Okeke, I. N. 2018 *Pathogenic Lineages of Enteric Bacteria in Nigeria, Grant Number: MR/L00464X/1.*

Eduvie, M. O., Olabode, T. & Yaya, O. O. 2003 Assessment of groundwater potentials of Abuja environs. In: *Proceedings of the 29th Water, Engineering and Development Centre-UK (WEDC) International Conference*, Held in Abuja-Nigeria, WEDC London. pp. 185–187.

Galway, L. P., Allen, D. M., Parkes, M. W., Li, L. & Takaro, T. K. 2015 Hydroclimatic variables and acute gastro-intestinal illness in British Columbia, Canada; a time series analysis. *Water Resour. Res.* **51**, 885–895. doi:10.1002/2014WR015519.

Hay, S. I., George, D. B., Moyes, C. L. & Brownstein, J. S. 2013 Bigdata opportunities for global infectious disease surveillance. *PLoS Med.* **10**, 1–4. doi:10.1371/journal.pmed.1001413.

Ibadan Rainfall (IR) 2019 *Monthly Weather Forecast and Climate Ibadan, Nigeria*. Available from: https://www.weather-nga.com/en/nigeria/ibadan-weather-january (accessed 10 September 2020).

Ince, M., Bashir, D., Oni, O. O. O., Awe, E. O., Ogbechie, V., Korve, K., Adeyinka, M. A., Olufolabo, A. A., Ofordu, F. & Kehinde, M. 2010 Rapid Assessment of Drinking water quality in the Federal Republic of Nigeria: In *Country Report of the Pilot Project Implementation in 2004*–2005. WHO Library Cataloguing-in-Publication Data. ISBN 978 92 4 150060 9.

Invik, J., Barkema, H. W., Massolo, A., Neumann, N. F. & Checkley, S. 2017 Total coliform and *Escherichia coli* contamination in rural well water: analysis for passive surveillance. *J. Water Health* **15** (5), 729–740.

Johnson, J. R., Murray, A. C., Gajewski, A., Sullivan, M., Snippes, P., Kuskowski, M. A. & Smith, K. E. 2003 Isolation and molecular characterization of nalidixic acid-resistant extraintestinal pathogenic *Escherichia coli* from retail chicken products. *Antimicrob. Agents Chem.* **47**, 2161–2168.

Kaper, J. B., Nataro, J. P. & Mobley, H. L. T. 2004 Pathogenic *Escherichia coli*. *Nat. Rev. Microbiol.* **2**, 123–140.

Kaplan, B. E. M., Jothikumar, N., Routh, J., Gómez, G. A., Mintz, E. D. & Hill, V. R. 2017 Environmental survey of drinking water sources in Kampala, Uganda, during a typhoid fever outbreak. *Appl. Environ. Microbiol.* **83**, 1706–1717. https://doi.org/10.1128/AEM.01706-17.

Lugli, G. A., Longhi, G., Mancabelli, L., Alessandri, G., Tarracchini, C., Fontana, F., Turroni, F., Milani, C., van Sinderen, D. & Ventura, M. 2022 Tap water as a natural vehicle for microorganisms shaping the human gut microbiome. *Environ. Microbio.* doi: 10.1111/1462-2920.15988. Advance online publication. https://doi.org/10.1111/1462-2920.15988.

Madec, J. Y., Haenni, M., Ponsin, C., Kieffer, N., Rion, E. & Gassilloud, B. 2016 Sequence type 48 *Escherichia coli* carrying the $bla_{CTX-M-1}$ inci1/ST3 plasmid in drinking water in France. *Antimicrobial* **60** (10), 6430–6432. https://doi.org/10.1128/AAC.01135-16.

Mancini, J., Weckselblatt, B., Chung, Y. K., Durante, J. C., Andelman, S., Glaubman, J., Dorff, J. D., Bhargava, S., Lijek, R. S., Unger, K. P. & Okeke, I. N. 2011 The heat-resistant agglutinin family includes a novel adhesin from enteroaggregative *Escherichia coli* strain 60A. *J. Bacteriol.* **193** (18), 4813–4820. https://doi.org/10.1128/JB.05142-11.

Matamoros, S., van Hattem, J. M., Arcilla, M. S., Willemse, N., Melles, D. C., Penders, J., Vinh, T. N., Hoa, N. T., Bootsma, M. C. J., van Genderen, P. J., Goorhuis, A., Grobusch, M., Molhoek, N., Lashof, A. L. O., Stobberingh, E. E., Verbrugh, H. A., de Jong, M. D. & Schultsz, C. 2017 Global phylogenetic analysis of *Escherichia coli* and plasmids carrying the mcr-1 gene indicates bacterial diversity but plasmid restriction. *Sci. Rep.* **7**, 15364 doi:10.1038/s41598-017-15539-7.

Molekoa, M. D., Avtar, R., Kumar, P., Thu-Minh, H. V. & Kurniawan, T. A. 2019 Hydrogeochemical assessment of groundwater quality of Mokopane area, Limpopo, South Africa using statistical approach. *Water*. doi:10.3390/w11091891. Available from: www.mdpi.com/journal/water.

Momtaz, H., Dehkordi, F. S., Rahimi, E. & Asgarifar, A. 2013 Detection of *Escherichia coli*, *Salmonella* species and *Vibrio cholerae* in tap water and bottled drinking water in Isfahan, Iran. *BMC Public Health* **13**, 556.

Nataro, J. P. & Kaper, J. B. 1998 Diarrhoeagenic *Escherichia coli*. *Clin. Microbiol. Rev.* **11**, 142–201.

National Population Commission (NPC) 2015 *Ibadan Population*. Available from: http://population.city/nigeria/Ibadan.

Nigerian Standard for Drinking Water Quality (NSDWQ) 2015 *Nigerian Industrial Standard NIS 554: 2007ICS 13.060.20: 1-30*.

North Dakota State University (NDSU) 2015 Variation in isolates of *Escherichia coli* obtained from surface soil. Under the following project numbers PRJNA416911 and SRP126391.

Nowicki, S., deLaurent, Z. R., de Villiers, E. P., Githinji, G. & Charles, K. J. 2021 The utility of *Escherichia coli* as a contamination indicator for rural drinking water: evidence from whole genome sequencing. *PLoS ONE* **16** (1), e0245910.

Okeke, I. N., Wallace-Gadsden, F., Simons, H. R., Matthews, N., Labar, A. S., Hwang, J. & Wain, J. 2010 Multi-Locus sequence typing of enteroaggregative *Escherichia coli* isolates from Nigerian children uncovers multiple lineages. *PLoS ONE* **5** (11), e14093. doi:10.1371/journal.pone.0014093.

Owusu-Edusei Jr., K.Jr. & Owens, C. J. 2009 Monitoring county-level chlamydia incidence in Texas, 2004–2005: application of empirical Bayesian smoothing and exploratory spatial data analysis (ESDA) methods. *Int. J. Health Geographics* **8**, 1–6. doi:10.1186/1476-072X-8-12.

Popoola, O., Kehinde, A., Ogunleye, V., Adewusi, O. J., Toy, T., Mogeni, O. D., Aroyewun, E. O., Agbi, S., Adekanmbi, O., Adepoju, A., Muyibi, S., Adebiyi, I., Elaturoti, O. O., Nwimo, C., Adeoti, H., Omotosho, T., Akinlabi, O. C., Adegoke, P. A., Adeyanju, O. A., Panzner, U., Baker, S., Par, S. E., Marks, F. & Okeke, I. N. 2019 Bacteremia among febrile patients attending selected healthcare facilities in Ibadan. *Nig. Clin. Infect. Dis.* **69** (S6), S466–S473.

Raosoft 2004 *Sample Size Calculator*. Available from: http://www.raosoft.com/samplesize.htmL

Rashid, M., Rakib, M. M. & Hasan, B. 2015 Antimicrobial-resistant and ESBL-producing *Escherichia coli* in different ecological niches in Bangladesh. *Infect.* **5**, 1. doi: 10.3402/iee.v5.26712.

Riley, L. W. 2014 Pandemic lineages of extraintestinal pathogenic *Escherichia coli*. *Clin. Microbiol. Infect.* **20**, 380–390. doi: 10.1111/1469-0691.12646.

Tarr, P. I., Bilge, S. S., Vary Jr., J. C.Jr., Jelacic, S., Habeeb, R. L., Ward, T. R., Baylor, M. R. & Besser, T. E. 2000 *Iha*: a novel Escherichia coli O157:H7 adherence-conferring molecule encoded on a recently acquired chromosomal island of conserved structure. *Infect Immun.* **68** (3), 1400–1407. https://doi.org/10.1128/IAI.68.3.1400-1407.2000.

Tetzschner, A. M. M., Johnson, J. R., Johnston, B. D., Lund, O. & Scheutz, F. 2020 In Silico genotyping of *Escherichia coli* isolates for extraintestinal virulence genes by use of whole-genome sequencing data. *J. Clin. Microbiol.* **58**, e01269–20.

United Nations 2013 *Water Security and the Global Water Agenda: A UN-Water Analytical Tool*. Available from: https://www.unwater.org/publications/water-security-global-water-agenda/.

United Nation Economic and Social Council 2017 Progress towards the sustainable development goals: report of the secretary-general. *Agenda Items* **5**, 6 and 18 (a) of 11 May 2017, pp. 8–9. N1713409 SDGs Update.

United States Department of Agriculture, Food Safety and Inspection Service 2020 Whole genome sequencing of cultured STECs as part of the GenomeTrakr surveillance project for rapid source tracking of foodborne illness outbreaks. GenomeTrakr Project: USDA – Food Safety and Inspection Service PRJNA268206 • SRP050539.

Wohlsen, T., Bates, J., Vesey, G., Robinson, W. A. & Katouli, M. 2006 Evaluation of the methods for enumerating coliform bacteria from water samples using precise reference standards. *Lett. Appl. Microbiol.* **42**, 350–356. doi:10.1111/j.1472-765X.2006.01854.x.

World Health Organization/United Nation Children Education Fund 2006 *Core Questions on Drinking-Water and Sanitation for Household Surveys*. WHO Press, World Health Organization, Geneva.

World Health Organization 2011 *Guidelines for Drinking-Water Quality*, 4th edn WHO Press, Geneva.

Zahra, R., Javeed, S., Malala, B., Babenko, D. & Toleman, M. A. 2018 Analysis of *Escherichia coli* STs and resistance mechanisms in sewage from Islamabad, Pakistan indicates a difference in *E. coli* carriage types between South Asia and Europe. *J. Antimicrob. Chemother.* **73**, 1781–1785. doi:10.1093/jac/dky109.

First received 14 May 2022; accepted in revised form 25 June 2022. Available online 8 July 2022