Journal Pre-proof

Development of a multiplex droplet digital PCR assay for simultaneous detection and quantification of Escherichia coli, E. marmotae, and E. Ruysiae in water samples

Marie Moinet, Rose M. Collis, Lynn Rogers, Megan L. Devane, Patrick J. Biggs, Rebecca Stott, Jonathan Marshall, Richard Muirhead, Adrian L. Cookson

PII: S0167-7012(24)00021-6

DOI: https://doi.org/10.1016/j.mimet.2024.106909

Reference: MIMET 106909

To appear in: *Journal of Microbiological Methods*

Received date: 3 December 2023

Revised date: 28 February 2024

Accepted date: 29 February 2024

Please cite this article as: M. Moinet, R.M. Collis, L. Rogers, et al., Development of a multiplex droplet digital PCR assay for simultaneous detection and quantification of Escherichia coli, E. marmotae, and E. Ruysiae in water samples, *Journal of Microbiological Methods* (2023), https://doi.org/10.1016/j.mimet.2024.106909

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2024 Published by Elsevier B.V.

Journal Pre-proof

Development of a multiplex droplet digital PCR assay for simultaneous detection and quantification of *Escherichia coli*, *E. marmotae,* and *E. ruysiae* in water samples

Marie Moinet [a]†, Rose M. Collis [a] *†, Lynn Rogers [a, b], Megan L. Devane [c], Patrick J. Biggs [b, d], Rebecca Stott [e], Jonathan Marshall [f], Richard Muirhead [g], and Adrian L. Cookson [a, b]*

[a] AgResearch Ltd., Food Systems Integrity Team, Hopkirk Research Institute, Tennent Drive, 4442 Palmerston North, New Zealand; Marie.Moinet@gmail.com (M.M.), Rose.Collis@AgResearch.co.nz (R.C.), Adrian.Cookson@AgResearch.co.nz (A.C.), Lynn.Rogers@agresearch.co.nz (L.R.)

[b] Massey University, *m*EpiLab, School of Veterinary Science, Hopkirk Research Institute, Tennent Drive, 4442 Palmerston North, New Zealand; P.Biggs@massey.ac.nz (P.B.)

[c] Institute of Environmental Science and Research Ltd. (ESR), 27 Creyke Rd, Ilam, 8041 Christchurch, New Zealand; Megan.Devane@esr.cri.nz (M.D.)

[d] Massey University, School of Natural Sciences, Tennent Drive, 4442 Palmerston North, New Zealand

[e] National Institute of Water and Atmospheric Research (NIWA), Gate 10 Silverdale Road, Hillcrest, 3216 Hamilton, New Zealand; Rebecca.Stott@niwa.co.nz (R.S.)

[f] Massey University, School of Mathematical and Computational Sciences, Tennent Drive, 4442 Palmerston North, New Zealand; J.C.Marshall@massey.ac.nz (J.M.)

[g] AgResearch Ltd., Ethical Agriculture, Invermay, 176 Puddle Alley, 9092 Mosgiel, New Zealand; Richard.Muirhead@AgResearch.co.nz (R.M.)

* Corresponding authors

† contributed equally to this work and share first authorship

Journal Pre-proof

Abstract

Escherichia coli are widely used by water quality managers as Fecal Indicator Bacteria, but current quantification methods do not differentiate them from benign, environmental *Escherichia* species such as *E. marmotae* (formerly named cryptic clade V) or *E. ruysiae* (cryptic clades III and IV). Reliable and specific techniques for their identification are required to avoid confounding microbial water quality assessments. To address this, a multiplex droplet digital PCR (ddPCR) assay targeting *lipB* (*E. coli* and *E. ruysiae*) and *bglC* (*E. marmotae*) was designed. The ddPCR performance was assessed using *in silico* analysis; genomic DNA from 40 local, international, and reference strains of target and non-target coliforms; and spiked water samples in a range relevant to water quality managers (1 to 1000 cells/100mL). Results were compared to an analogous quantitative PCR (qPCR) and the Colilert method. Both PCR assays showed excellent sensitivity with a limit of detection of 0.05pg/µL and 0.005pg/µl for ddPCR and qPCR respectively, and of quantification of 0.5pg/µL of genomic DNA. The ddPCR allowed differentiation and quantification of three *Escherichia* species per run by amplitude multiplexing and showed a high concordance with concentrations measured by Colilert once proportional bias was accounted for. *In silico* specificity testing underlined the possibility to further detect and distinguish *Escherichia* cryptic clade VI. Finally, the applicability of the ddPCR was successfully tested on environmental water samples where *E. marmotae* and *E. ruysiae* potentially confound *E. coli* counts based on the Most Probable Number method, highlighting the utility of this novel ddPCR as an efficient and rapid discriminatory test to improve water quality assessments.

Keywords

absolute quantification, digital PCR, Fecal indicator organisms, naturalized *Escherichia*, quantitation methods, water microbial quality

Journal Pre-proof

Abbreviations[*]

1 Introduction

Escherichia coli are present in high concentrations in the gut and feces of mammals and birds and have been widely used as fecal indicator bacteria to monitor freshwater quality [1]. The relatively good correlation between the number of *E. coli* and the presence of pathogenic bacteria, protozoa, or viruses from fecal sources in water is used to determine safe levels for drinking or recreational water [2].

E. coli quantification is most commonly undertaken by membrane filtration or the 'most probable number' (MPN) method. For the latter, commercially available kits like the Colilert/Quantitray (IDEXX) system make the culture and enumeration relatively simple and convenient with reduced sample handling and processing and are widely used by water quality managers. However, these methods rely on the ability of *E. coli* to utilize specific growth substrates and are prone to sensitivity and specificity issues [3, 4], for instance enterohaemorrhagic *E. coli* O157:H7 lack the β-glucuronidase activity on which Colilert tests are based [5]. Also, recently described 'cryptic clades' (II to VIII) of *Escherichia* are phenotypically indistinguishable from fecal *E. coli* [6-8]. They include three formally described novel species, *Escherichia whittamii* sp. nov. for clade II [9], *Escherichia marmotae* for clade V [10] and *Escherichia ruysiae* for clades III and IV [11]. Whilst there have been a small number of reports to suggest that *E. marmotae* and *E. ruysiae* can be isolated from humans and may become systemic and associated with bacteriaemia [6, 12] evidence from New Zealand indicates they are mostly non-pathogenic and more prevalent from wildlife and environmental samples (water, sediment, periphyton) [13-15]. The cryptic clades exhibit β-glucuronidase activity and are also detected by enumeration methods such as the Colilert/Quantitray systems, potentially leading to

Abbreviations: *bglC*: 6-phospho-beta-glucosidase gene, Cq: Cycle threshold, ddPCR: droplet digital Polymerase Chain Reaction, eDNA: environmental DNA, gDNA: genomic DNA, *lipB*: lipoate-protein ligase B gene, LOD: Limit of detection, LOQ: Limit of Quantitation, MPN: Most Probable Number, NTC: No Template Control, NZ: New Zealand, qPCR: quantitative real-time PCR, SBA: Sheep Blood Agar

Journal Pre-proof

confounded water quality assessments [16]. The New Zealand freshwater policy [17] describes critical *E. coli* concentrations for action e.g. to focus on samples requiring regulatory authority action in New Zealand: (i) 260 *E. coli* /100 mL to spur water managers to identify sources of fecal contamination for mitigation purposes; and (ii) 540 *E. coli* /100 mL to inform the public that the site is unsuitable for primary contact recreational use. Their presence in water samples could result in the erroneous notification of a public health risk when none exists with identification and quantification of *E. marmotae* and *E. ruysiae* required when Colilert MPN data are >260 to 540 MPN/100 mL, and no obvious faecal sources can be identified.

Previous studies have identified *E. marmotae* and to a much lower extent *E. ruysiae* as the most prevalent *Escherichia* clades detected in the environment in New Zealand [13]. They are also the most common cryptic clades isolated worldwide from bird feces and aquatic environments [9, 18-20], although there are exceptions with clades I and II found at higher prevalence [21, 22].

Current methods to differentiate enteric *E. coli* from environmental *Escherichia* are mostly culture dependent. To date, no phenotypic tests have been able to differentiate cryptic *Escherichia* species from *E. coli* 'sensu stricto' [23-25], and only genotypic (molecular) methods have been used. After putative *E. coli* have been isolated, distinction of the species can be achieved using the 'Clermont PCR' [24, 26], by 'eccPCR' (the *ecc* gene is found in cryptic clades II-VIII, but not in *E. coli* [27]), or by PCR targeting the *ycjM* (glucosyltransferase) gene (in most *E. coli* but only 23% of the environmental strains [28]). When high *E. coli* counts are observed in waterways with no clear resolution of likely fecal sources, distinguishing *E. coli* from cryptic clades can take time and resources [29]. There is evidence suggesting that 'naturalized' *E. coli*, of fecal origin and able to survive wastewater treatment, also carry numerous resistance and virulence factors [30, 31], and thus present a public health risk. In addition, naturalized *E. coli* signal the potential presence of other fecal pathogens (e.g., *Giardia*) that also have the ability to persist in aquatic environments. New tools are required to distinguish benign, environmental *Escherichia* from fecal *E. coli* and to provide confidence for water

Journal Pre-proof

quality managers ensuring appropriate mitigations are implemented directed at reducing *E. coli* derived from fecal sources.

Droplet digital PCR (ddPCR) is a highly precise nucleic acid quantification method that partitions a sample into thousands of nanoliter-sized droplets, allowing for absolute quantification with high tolerance to PCR inhibitors [32, 33]. It is, therefore, a promising tool for *Escherichia* detection and quantification in environmental samples. So far, ddPCR methods targeting *Escherichia* have been developed with a focus on pathogenic *E. coli* and mostly in food matrices [34-39]. In this article, we described the development of a ddPCR for simultaneous detection and quantification of *E. coli, E. ruysiae* and *E. marmotae* from water samples. We also compared the detection performance of this ddPCR assay against an analogous qPCR assay and tested its applicability on environmental water samples.

2 Material and methods

2.1 Reference / overseas strains

The local and overseas strains used for the development and validation of the PCR assays in this study are listed in Table 1. A panel of strains belonging to the target species and representative of the populations of *Escherichia* spp. found both in New Zealand and internationally was created, as well as a reference panel of other non-target coliforms. The strains of *E. marmotae*, *E. ruysiae* and *E. coli* used for the development of the PCR assays in this study were AGR4200, AGR4111 and AGR4274, respectively [13]. For these three strains, cultures were grown on Sheep Blood Agar (SBA, Fort Richard, New Zealand) and a single colony was inoculated in EC broth, (Oxoid, Hampshire, UK) incubated at 35°C for 18 to 21 hours, and subsequently used to prepare a genomic DNA dilution series (see section 2.4.1 below). Other non-target strains used for the PCR assay validation were grown on SBA and a single colony inoculated in Brain Heart Infusion broth (Oxoid, Fort Richard, New Zealand) and incubated at 37°C for 18 to 21 hours, except for *Bacillus cereus* (30°C for 18 to 21

Journal Pre-proof

Journal Pre-proof

hours) and *Serratia liquefaciens* and *Hafnia alvei* (25°C for 48 hours). All strains were cultivated aerobically.

Table 1: Local and reference strains of target and non-target species used in this study.

Species	Strain	Prov.[a]	Synonyms	Accession No.[b]	*lipB*[c]	*bglC*
E. coli Phylogroup A	AGR4148	[13]	.	SAMN12996513	+	-
E. coli phylogroup B1	AGR4274	[13]	.	SAMN12996551	+	-
E. coli Phylogroup B2	AGR3711	[13]	.	SAMN12996373	+	-
E. coli Phylogroup C	AGR4288	[13]	.	SAMN12996556	+	-
E. coli Phylogroup D	AGR4073	[13]	.	SAMN12996492	+	-
E. coli Phylogroup E	AGR3739	[13]	.	SAMN12996381	+	-
E. coli Phylogroup F	AGR4155	[13]	.	SAMN12996515	+	-
E. coli	NZRM 916	NZRM	ATCC 25922	ATCC genomes	+	-
E. coli O157:H7	NCTC12900	NCTC	ATCC 700728	ATCC genomes	+	-
E. marmotae	AGR4200	[13]	.	SAMN12996534	-	+
E. marmotae	DSM28771^{T}	DSMZ	HT073016^{T}	CP025979	-	+
E. marmotae	NCTC 11133	NCTC	.	LR134340	-	+
E. marmotae	NZRM29	NZRM	NCTC8196	LR134270	-	+
E. marmotae	MOD1-EC5110	FDA	.	PTTC01000000	-	+
E. marmotae	MOD1-EC5427	FDA	.	PTTA01000000	-	+
E. marmotae	MOD1-EC5438	FDA	.	PTSZ01000000	-	+
E. marmotae	MOD1-EC5462	FDA	.	PTSW01000000	-	+
E. marmotae	MOD1-EC5948	FDA	.	PTST01000000	-	+
E. marmotae	MOD1-EC5949	FDA	.	PTSS01000000	-	+
E. marmotae	MOD1-EC6150	FDA	.	PTSG01000000	-	+
E. marmotae	MOD1-EC6163	FDA	.	PTSA01000000	-	+
E. ruysiae	AGR4111	[13]	.	SAMN12996506	<+	-
E. ruysiae	NCTC 14359^{T}	NCTC	OPT1704	CABVLQ01	<+	-
E. ruysiae	MOD1-EC6819	FDA	.	PTRG01000000	<+	-
E. ruysiae	MOD1-EC6842	FDA	.	PTRF01000000	<+	-
E. fergusonii	ATCC35469	ATCC	.	ATCC genomes	-	-
E. fergusonii	ATCC35470	ATCC	.	.	-	-
E. fergusonii	ATCC35472	ATCC	.	.	-	-
Bacillus cereus	NZRM5	NZRM	ATCC 10702	ATCC genomes	-	-
Citrobacter freundii	NZRM 982	NZRM	ATCC 8090	ATCC genomes	-	-
Hafnia alvei	NZRM 860	NZRM	CDC 5190-70	.	-	-
Klebsiella aerogenes	NZRM 798	NZRM	ATCC 13048	ATCC genomes	-	-
Klebsiella oxytoca	NZRM 2301	NZRM	ATCC 13183	ATCC genomes	-	-
Klebsiella pneumoniae	NZRM3681	NZRM	ATCC 700603	ATCC genomes	-	-
Rahnella aquatilis	NZRM 3285	NZRM	.	.	-	-
Raoultella ornithinolytica	DF_0025a	[40]	.	.	-	-

Journal Pre-proof

Salmonella enteritidis	NZRM 351	NZRM	ACTT 13076	ATCC genomes	-	-
S. typhimurium	LT2	unk.	.	NC_003197	-	-
Serratia liquefaciens	NZRM 2585	NZRM	ATCC 27592	ATCC genomes	-	-
Staphylococcus aureus	NZRM 917	NZRM	ATCC 25923	ATCC genomes	-	-

[a] Provenance: NZRM, New Zealand Reference Culture Collection: Medical Section; NCTC, UK National Collection of Type Cultures; DSMZ, German Collection of Microorganisms and Cell Cultures; ATCC, American Type Culture Collection; FDA, US Food and Drug Administration; [b] ATCC genomes available at https://genomes.atcc.org ; [c] "<+" denotes a positive ddPCR amplification at a lower amplitude; [T] = type strain

2.2 Spiked samples

To evaluate the performance of the qPCR and ddPCR assays in environmental water samples and compare them with the Colilert method, a known concentration of *E. coli* (AGR4274) and *E. marmotae* (AGR4200) was serially diluted and spiked in water prior to DNA extraction. First, the cultures were grown on SBA and a single colony was inoculated in 10 mL M9 minimal medium [41] at 35°C overnight (18 hours). The dilution series was then made in two steps to specifically encompass the range of concentrations of interest for water quality managers, *i.e.,* zero to thousands of cells per 100 mL of water, based on the assumption that the concentration of an overnight culture is approximately 1 x 10^9 cells/mL. The overnight culture was first diluted in phosphate buffered saline (PBS, 0.01M, pH7.2) to a starting concentration of ~2.2 x10^4 cells/mL (Dilution A), which was then serially diluted with a 1:4 ratio (Dilutions B to F, Table 2). Each dilution was then spiked (1 mL) in 2.2 L of sterile Reverse Osmosis (RO) water. From this suspension, 100 mL was quantified using the Colilert-18 Quanti-Tray/2000 method as per the manufacturer's instructions, 2 L was used for DNA extraction (see 2.4.2), and 100 µL after a 1:3 dilution (for dilution A) or 200 µL (for dilutions B-F) were plated onto SBA and incubated at 35°C overnight for viable counts [42].

Table 2: Expected concentrations (Con.) for dilution series A to F and subsequent assays starting from a hypothetical overnight culture concentration of 1 x 10^9 cells/mL. The 2.2 L in which the initial dilution was spiked was used for Colilert quantification (100 mL), DNA extraction (2L), and viable count on plates (100 to 200 µL).

Dilution series	Initial con. (cells/mL)	Spiked volume (mL)	Con. in 2.2L (cells/mL)	Cells/100 mL (Colilert)	Cells/2L (DNA extract.)	Expected colonies in 200µL	Expected colonies in 100µL (1:3 diln)
A	2.2 x 10^4	1	10	1000	20000	ND*	367
B	5500	1	2.5	250	5000	275	-

C	1375	1	0.625	62.5	1250	69	-
D	343.2	1	0.156	15.6	312	18	-
E	85.8	1	0.039	3.9	78	5	-
F	21.45	1	0.00975	0.975	19.75	2	-

*Not Done

2.3 Environmental water samples

In addition to spiked water samples, environmental DNA (eDNA) was extracted from three water samples and run through the ddPCR assay. Water was collected as previously described [13] from the inflow (W1), middle (W2), and outlet cells (W3) of a constructed wetland in the Toenepi River catchment, Waikato, New Zealand. This constructed wetland receives inflow from surrounding dairy farm tile drainage systems, with consistently higher *E. coli* concentrations measured in the outflow compared to the inflows [29]. Investigations into the underlying causes of this apparent poor performance in treating agricultural run-off have pinpointed the persistence and growth of naturalized *E. coli* phylogroup B1 in the wetland, but *E. marmotae* and *E. ruysiae* have also been isolated there [13, 29]. These findings, therefore, provide an ideal real-life scenario to assess the practical applicability of the ddPCR assay developed in this article. Water samples were stored at +4°C until arrival at the laboratory and processed the following day. The Colilert-18 Quanti-Tray/2000® method (IDEXX, NZ) was used as above for the enumeration of coliforms and generic '*E. coli*' per 100 mL of water, in parallel to eDNA extraction (see section 2.4.2 below). The eDNA was stored at -20°C and tested 26 days after sampling.

2.4 DNA extraction

2.4.1 From pure cultures

Microbial DNA was extracted from 1 mL of pure culture using the Wizard® Genomic DNA Purification Kit (Promega, USA) following the manufacturer's instructions for Gram-positive (*Bacillus cereus* and *Staphylococcus aureus*) or Gram-negative bacteria (all others), with minor modifications as the DNA pellet was rehydrated using 50 µL Tris-HCl (pH 8.0). The genomic DNA (gDNA) concentration was measured with a Qubit™ 4 Fluorometer (Thermo Fisher Scientific, USA) using the dsDNA BR Assay Kit

Journal Pre-proof

(Thermo Fisher Scientific, USA) and the gDNA purity and presence of contaminants was checked with the A260/280 and A260/230 ratios using a Nanodrop 2000c spectrophotometer (Thermo Fisher Scientific, USA).

2.4.2 From environmental and spiked samples

DNA was extracted from 1 L of environmental water samples and from 2 L of spiked RO water using the DNeasy® PowerSoil® Pro kit (Qiagen, NZ) with the manufacturer's instructions slightly modified. Water was filtered within 24 hours using positive pressure on a 0.45 µm nitrocellulose filter (diameter 47mm; S-Pak Membrane Filters, Millipore) and the filter (or filters when the turbidity was too high to filter all the volume at once) placed in a 7 mL tube with 800 µL of CD1 lysis buffer and stored at -20°C until further processed. Tubes, containing the filter and lysis buffer, were thawed on ice, zirconium beads from a PowerBead Pro tube added and bead beating done with a Mini-BeadBeater 16 (BioSpec Products, OK, USA) and two 30 second beating durations with a one minute interval in between. The remaining Quick-Start protocol was followed according to the manufacturer's instructions, except that the DNA was eluted in 30 µL of Solution C6. The DNA was further purified to improve A260/280 and A260/230 quality control attributes using sodium chloride and ethanol precipitation as per the manufacturer's instructions and the DNA resuspended in a final volume of 50 µL of Solution C6. The DNA concentration and the presence of contaminants were checked as described above. Two liters of sterile RO water were filtered and processed similarly, and the extraction blank obtained used in PCRs as a No Template Control (NTC).

2.5 Preparation of gDNA dilution series

To study the dynamic range and sensitivity of the PCR assays, a dilution series was prepared using the stock solution of gDNA extracted from AGR4200, AGR4111 and AGR4274 for *E. marmotae*, *E. ruysiae* and *E. coli*, respectively. The gDNA concentration previously measured with Qubit™ was diluted to a working concentration of 5 ng/µL (*i.e.,* roughly 10^6 genome equivalent) in Ambion nuclease-free water (Invitrogen, Thermo Fisher Scientific, USA) and a 10-fold serial dilution series

Journal Pre-proof

ranging DNA concentrations of 5 ng/µL to 5 fg/µL (*i.e.* over 7 orders of magnitude) was prepared using nuclease-free water.

2.6 Primers and probes design

Pangenome analysis data from *E. coli*, *E. ruysiae* and *E. marmotae* isolates representative of different locations in New Zealand obtained in a previous study [13] was used to find candidate assay targets. The single-copy lipoate-protein ligase B gene (*lipB*) was chosen to identify *E. coli* and *E. ruysiae*, and the single-copy 6-phospho-beta-glucosidase *bglC* gene was chosen to identify *E. marmotae.* The visualization and design of specific target primers and 5'-FAM and 5'-HEX TaqMan™ probes were performed using Geneious Prime 2022.2.1. Designed primers and probes (Table 3) were synthetized by LGC BioSearch technologies (Petaluma, CA, USA) and used in both qPCR and ddPCR assays. Specificity of the primers was checked *in silico* using NCBI Primer-BLAST tool (https://www.ncbi.nlm.nih.gov/tools/primer-blast/) with the non-redundant (nr) NCBI database, using the default parameters except the exclusion of uncultured/environmental samples, and checking only for bacterial organisms. Some strains of *E. marmotae* previously submitted to sequence databases were recorded as *E. coli*, therefore strain accessions matching the *bglC* primers were further checked using the online 'ClermonTyper' tool [43]. In addition, the primers and probes were mapped to available reference genomes (Table 1) and draft genomes of cryptic clade VI (IDs MOD1-EC7003, UHCL_3L, and UHCL_L1), clade VII (ROAR019) and clade VIII (250f3330-fa7c-1, FSIS12033615, FSIS12033877, FSIS12036872, FSIS12037041, PSU-2607, and SCPM-O-B-8794) downloaded from EnteroBase [8], using Geneious Prime 2022.2.1, allowing up to 2 mismatches in the binding region.

Table 3: lipB and bglC primers and probes used in this study for Escherichia coli and E. marmotae PCR amplifications (respectively).

Designation	Sequence (5'-3')	Target	Amplicon Size
lipB probe	6-FAM-TGA TGA AAT CTG GCT GGT CGA GCA CTA TCC-BHQ1	*lipB*	133 bp
lipB-F	CAG CCT TAC GAG CCA ATC TCC C		
lipB-R2	TGT GCT CCG CYT TTC CTG CCT G		

*bglC*2 probe	560/HEX- CGT TTG TTC GTT ACG CCC GTC TGC TGT T -BHQ1	*bglC*	127 bp
*bglC*2-F	TGT ATC ATT TCG ACT TGC CCT G		
*bglC*2-R	CAT AGC GGG ACT TTA TCG GCA A		

6-FAM = 6-carboxyfluorescein; BHQ-1 = Black Hole Quencher; 560/HEX = hexachlorofluorescein

2.7 qPCR assay

The qPCR assay was performed on a Qiagen Rotor-Gene Q machine (Bio-Strategy Ltd, Auckland, NZ) using the following cycling conditions: initial denaturation at 95°C for 5 minutes, followed by 40 cycles (to establish limits of detection with purified genomic DNA) or 35 cycles (using bacterial DNA extracted from spiked water samples to enumerate bacteria) denaturation at 95°C for 10 seconds and annealing/extension at 60°C for 20 seconds. Each reaction contained 4 µL of ToughMix® (Quantabio), 0.8 µL of each primer at a final concentration of 250 nM, 0.3 µL of each probe at a final concentration of 100 nM, 10.2 µL of PCR grade water, and 2 µL of DNA template in a total volume of 20 µL. Each sample was run at least in triplicate. No optimization of annealing temperature or primers-probes concentration was needed. The raw data was analyzed in Rotor-Gene Q Series software version 2.3.1 to obtain a quantification cycle (Cq) value.

2.8 ddPCR assay

The ddPCR reaction mixture consisted of 11 µL of 2× ddPCR Supermix for Probes (no dUTP) (#1863024, Bio-Rad, NZ), 1.1 µL each of 20 × *lipB* or *bglC* probe and primers stock solution (0.25 µM final concentration for each probe and 0.9 µM final concentration for each primer), 2.2 µL of template DNA, and nuclease-free water to a final volume of 22 µL. To limit pipetting errors, 20 µL only of the mixture was loaded in a DG8TM cartridge and droplets generated using the QX200 Droplet Generator (Bio-Rad, NZ). The droplets were transferred to a 96-well PCR plate and PCR was performed using a C1000 Touch Thermal Cycler (Bio-Rad, NZ). No optimization of primers or probes concentrations was needed. The optimal annealing temperature was determined by performing a temperature gradient (55° - 65°C) within a single run using the built-in option of the thermal cycler, with pure culture DNA of *E. coli*, *E. marmotae* or *E. ruysiae* in simplex, duplex (*E. coli* + *E. marmotae*) or triplex. The final (optimized) cycling conditions were as follows: 10 min at 95°C for Taq activation,

Journal Pre-proof

followed by 40 cycles of denaturation at 94°C for 30 seconds and annealing/extension at 62°C for 1 minute, then a final enzyme deactivation step at 98°C for 10 minutes. Positive droplets were counted in a QX200 Droplet Reader (Bio-Rad, NZ) and raw data analyzed using QX Manager Standard Edition version 1.2.34530909. Thresholding was done manually according to the manufacturer's recommendations to obtain a total number of gene copies per µL. This number was then converted into gene copies per 100 mL environmental or spiked water sample using the following equation:

$$\text{Copies per 100 mL} = \text{Copies per µL} \times (\text{Total ddPCR reaction Volume} \div \text{Volume of DNA per reaction}) \times \text{Volume in µL of DNA extract obtained} \times (\text{Volume in mL of original sample} \div 100\text{ mL})$$

Each sample was run in triplicate. NTC and positive controls were added to each run and samples with less than 10,000 droplets generated were excluded from analysis and repeated until they passed this threshold.

2.9 Assays performance assessment

The limit of detection (LOD) of the qPCR and ddPCR assays, defined as the lowest concentration of nucleic acid or number of cells which gives a positive PCR result in over 80% of all replicates tested, was assessed using the gDNA dilution series described above. The limit of quantification (LOQ) was defined as the lowest standard concentration that could be quantified with a coefficient of variation (CV) value below 25%. The discrete threshold approach method and associated R script developed for the determination of LOD and LOQ for qPCR assays [44] was used and adapted for ddPCR outputs. Six replicates for each dilution were used. The LOD was also estimated using the dilution series of spiked water samples. For qPCR the Cq values were recalculated to the number of copies/µL before calculation of the CV, as recommended [45], while the ddPCR results were already expressed in copies/µL and the CV calculated directly. To estimate the repeatability (intra-assay precision) of the ddPCR method, sample concentrations from replicates were recorded individually rather than samples merged in the QX Management software and the CV calculated. Agreement between Colilert and ddPCR results for the dilution series of spiked water samples was assessed

Journal Pre-proof

using Bland and Altman's method [46]. Analyses and visualizations were done in R 4.2.2 using *ggplot2* 3.4.1 [47].

3 Results and discussion

3.1 Primers and probes design and *in silico* specificity

The specificity of both *lipB* and *bglC* assays tested *in silico* was excellent. The *lipB* sequences in *E. ruysiae* had no mismatch on the forward primer, one mismatch on the probe and two mismatches on the reverse primer (Figure 1). These base mismatches impacted the amplitude of *lipB* amplification for *E. ruysiae* compared to *E. coli* and were taken advantage of for ddPCR direct quantification of three targets per well by amplitude multiplexing. *In silico* primer and probe mapping on newly described clades VI-VIII genomes showed that amplification of *lipB* for clade VI strains was also possible (no mismatch on the forward primer or probe and two mismatches on the reverse primer), but the level of amplification relative to *E. coli* and *E. ruysiae* and the possibility of differentiating them with the ddPCR assay developed in this article needs to be confirmed with *in situ* testing. The primer-BLAST tool used to test each assay *in silico* did not reveal identical sequences other than those targeted except for strains of *Shigella* spp. for *lipB* (which was expected) and for three strains of *E. coli* for *bglC*, namely MSB1_5C-sc-2280313, ATCC 700415, and UPEC132 (accession numbers LR890576.1, CP022609.1, and CP040456.1, respectively, Figure 2). For the latter, the ClermonTyping results indicated the two first strains are misidentified *E. marmotae* strains (profile ----), as was already pinpointed by DNA-DNA hybridization for MSB1_5C-sc-2280313 [27]. The *E. coli* strain UPEC132 was an intriguing strain, it was Clermont-typed as belonging to phylogroup B2 (profile -++-) and also matched with the *lipB* primers and probe (with one mismatch on *lipB*-R2), the only strain of *E. coli* to our knowledge to do so. The primer-BLAST tool also identified partial matches for *bglC* primers with sequences of *Klebsiella* sp., *K. huaxiensis*, *Dickeya parazeae* and *D. aquatica* with three mismatches on at least one primer, and three (*D. parazeae*) to five or more (all others) mismatches on the probe (Figure 2). While the number of mismatches was relatively high, especially

Journal Pre-proof

on the probe, and unlikely to lead to an efficient detection, only reference strains of *Klebsiella* were included in the testing panel (see section 3.2), and phytopathogens of the genus *Dickeya* that have been isolated in low frequency from waterways [48-50] could be tested in future assays to ascertain the absence of non-specific amplification. The *bglC* primers and probe targeting *E. marmotae* did not match any of the other cryptic clades, making *bglC* an interesting target to differentiate clade V strains from other cryptic clades.

Figure 1: Alignment of the partial lipB *gene locus of targeted* Escherichia coli *and* E. ruysiae *sequences and selected non-targeted reference sequences illustrating the position of the forward primer (dark green), probe (red) and reverse primer (light green) for the lipB qPCR and ddPCR assays developed in this study.*

Journal Pre-proof

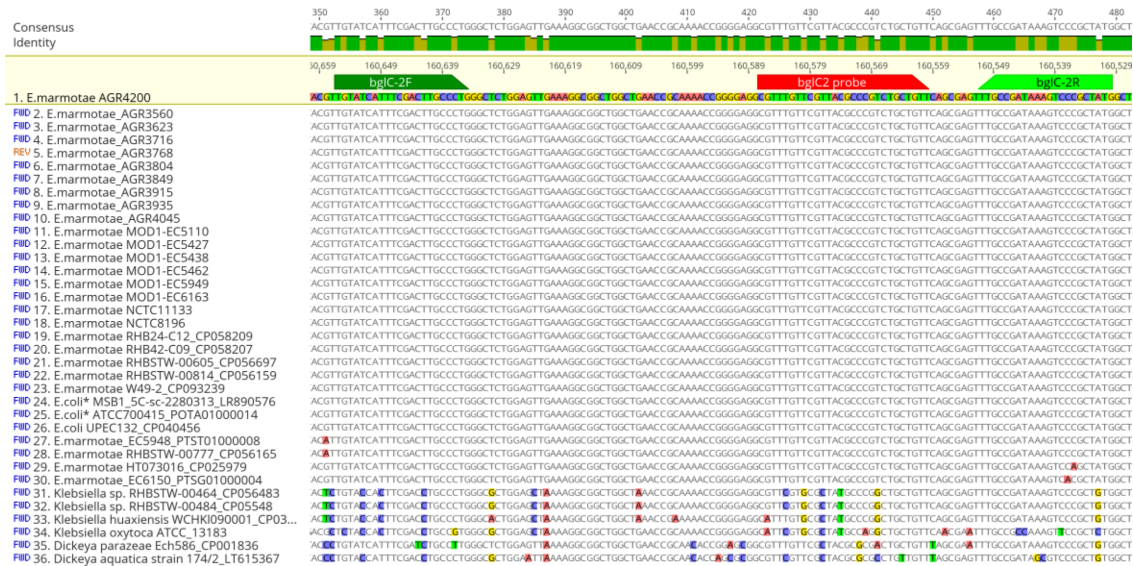


Figure 2: Alignment of the partial bglC *gene locus of targeted* Escherichia marmotae *sequences and partial matches identified by Primer-BLAST illustrating the position of the forward primer (dark green), probe (red) and reverse primer (light green) for the* bglC *qPCR and ddPCR assays developed in this study. * denotes an isolate of* E. marmotae *incorrectly classified (see main text). Sequences from NZ isolates of* E. marmotae *strains AGR3560 to AGR4045 come from the BioProject PRJNA576546 [13].*

3.2 Specificity

All target and non-target species tested with the ddPCR in this study gave anticipated results, including the *Klebsiella* strains (Table 1). As primers and probes were similar for the ddPCR and qPCR, the specificity was tested only with one method. The ddPCR 2D-amplification profile of pure cultures of each species alone or in combination confirmed the possibility to distinguish and quantify *E. marmotae*, *E. coli* and *E. ruysiae* (Figure 3). This is where the ddPCR method showed a clear advantage over the qPCR method, as the distinction between *E. coli* and *E. ruysiae* is not possible with the latter. According to the temperature gradient run conducted, an annealing temperature of 62°C was chosen for the ddPCR to optimize the differentiation between *E. coli* and *E. ruysiae*

amplification. Two strains of *E. marmotae* (DSM28771 and EC6150) presented a single mismatch on the *bglC*-2R primer *in silico* (Figure 2) that did not impact correct amplification with the ddPCR.

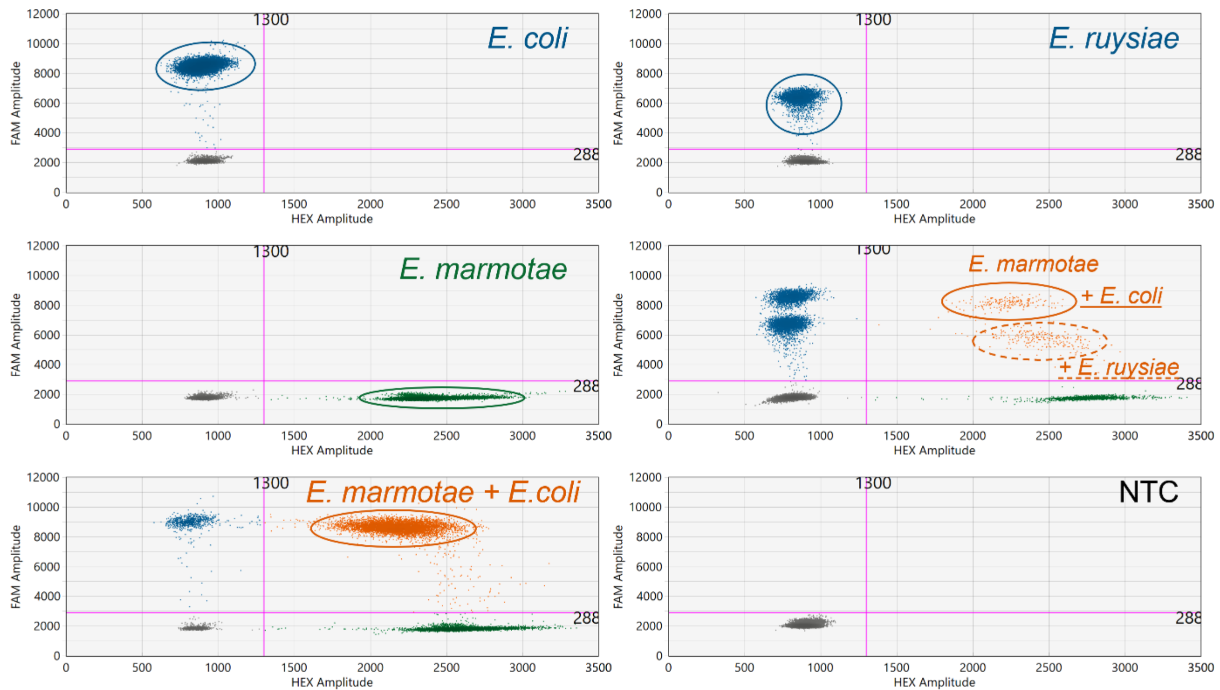


Figure 3: lipB (in blue) and bglC (in green) ddPCR 2D-amplification profiles of pure cultures of the target species Escherichia coli*,* E. ruysiae *and* E. marmotae *alone or in combination. NTC: blank extraction No Template Control. Orange points represent droplets positive for both targets (*E. marmotae *for* bglC *+ E. coli or E.ruysiae for* lipB*).*

3.3 Sensitivity

3.3.1 Reference strains gDNA

The sensitivity of both ddPCR and qPCR assays was excellent (Table 4A). The LOD was 0.05 pg/µL for both targets by ddPCR and 0.005 pg/µL for both targets by qPCR, meaning the assays were able to detect in the order of one target copy/µL. The LOQ was 0.5 pg/µL for *bglC* by qPCR and for both targets by ddPCR, and 0.05 pg/µL for lipB by qPCR, with a good intra-assay repeatability. The CV ranged between 11% and 24% for qPCR and between 11% and 14% for ddPCR at the LOQ concentration. The lower amplitude of amplification for *E. ruysiae* did not impact the LOD (determined by ddPCR only) that was similar to the LOD for *E. coli* (Table 4A), and the LOQ calculated

was lower, with a CV of 20% at 0.05 pg/µL. As for other eDNA assays, results below these LOQ and LOD should be interpreted qualitatively, as detections or non-detections [44].

3.3.2 Spiked water samples

The LOD in spiked water samples for both the ddPCR (15.6 cells/100 mL for *bglC* and 62.5 cells/100 mL for *lipB*) and qPCR assays (250 cells/100 mL for both targets, Table 4B), was below the range chosen to represent concentrations of interest for water quality managers, although the qPCR had a lower number of cycles (n = 35) compared to the ddPCR (n = 40). Different jurisdictions set different acceptable thresholds for fecal indicator bacteria in freshwater. For instance, the alert threshold is set at 260 *E. coli* /100 mL in the New Zealand national recreational water quality guidelines [17], and many countries use the threshold of 100 *E. coli*/100 mL or above [51]. The ddPCR assay described in this study is therefore appropriate to use when water quality managers are required to investigate the potential presence of environmental *Escherichia* during *E. coli* exceedances. At sites where *E. ruysiae* is not expected to be present in high concentrations, the more broadly available and less costly qPCR method could also be useful when *E. coli* counts are ≥ 250 *E. coli*/100 mL, with the inclusion of external DNA standards, however, required for absolute quantification.

Journal Pre-proof

Table 4: Sensitivity and performance of the ddPCR and qPCR assays obtained from (A) tenfold dilution series of Escherichia coli *(AGR4274),* E. marmotae *(AGR4200) and* E. ruysiae *(AGR4111) genomic DNA and (B) water samples artificially spiked with four-fold dilutions of* E. marmotae *and* E. coli*.*

A) Pure Culture genomic DNA																			
target	strai n	Dilution 1 5000 pg/µL		Dilutio n 2 500 pg/µL		Dilutio n 3 50 pg/µL		Dilutio n 4 5 pg/µL		Dilutio n 5 0.5 pg/µL		Dilutio n 6 0.05 pg/µL		Dilutio n 7 0.005 pg/µL		Dilutio n 8 0.0005 pg/µL		linear regression y=	R^2
								qPCR Average Ct ± sd (6 replicates)											
lipB	AGR 427 4	13 .7 5	± 0. 38	1 7. 6	± 0. 3	2 1. 0	± 0 .	2 4. 8	± 0. 1	2 8 .	± 0. 3	3 2 .	± 0 .	3 6. 4	± 1. 2	3 6. 9	\	- 3.67 7x +	0. 99 95

				7	8	6	21	3	4	45	0	19	41	5	9	3[a]		27.36	
bglC	AGR 4200	14.09	± 0.12	17.31	± 0.19	20.64	± 0.15	24.12	± 0.14	27.16	± 0.15	30.21	± 0.63	34.27	± 1.36	34.29[a]	\	-3.286x + 26.22	0.999
								ddPCR Average copies/µL ± sd (6 replicates)											
lipB	AGR 4274	*	\	7205.7	± 580.4	924.1	± 91.0	89.3	± 4.0	8.1	± 0.9	0.9	± 0.5	0.1[c]	± 0.0	ND	\	1.015 x + 1.22	0.9989
lipB	AGR 4111	**	\	8538.5	± 315.9	775.2	± 73.3	69.1	± 6.0	7.2	± 0.4	0.6	± 0.1	0.1[d]	± 0.0	ND	\	1.013 x + 1.15	0.9977
bglC	AGR 4200	*	\	5751.0	± 111.7	533.1	± 21.6	50.6	± 4.5	5.1	± 0.7	0.6[b]	± 0.2	0.1[d]	± 0.0	ND	\	0.99x + 1.04	0.9984

* = saturated signal ; ** = saturated signal at lower amplitude; ND = Not detected; \ = not calculated; [a] not detected in 5 duplicates; [b] not detected in 1 replicate; [c] not detected in 4 replicates; [d] not detected in 3 replicates

B) Spiked water samples

target	strain	Dilution A 1000 cells/100mL		Dilution B 250 cells/100mL		Dilution C 62.5 cells/100mL		Dilution D 15.6 cells/100mL		Dilution E 3.9 cells/100mL		Dilution F 0.975 cells/100mL	
						Quantitray							
		MPN	95%CI	MPN	95%CI	MPN	95%CI	MPN	95%CI	MPN	95%CI	MPN	95% CI
	AGR 4274	1046.2	[705-1509]	248.1	[162.3-371.9]	78.9	[56.3-108.8]	20.1	[12.4-31.8]	4.1	[1.7-9.5]	1	[0.1-5.5]
	AGR 4200	1553.1	[1016.2-2353.1]	325.5	[206.6-498.1]	95.9	[68.4-130.5]	9.8	[4.7-18.4]	7.5	[3.6-14.9]	<1.0	[0-3.7]
						Actual concentrations (from viable counts)							
	AGR 4274	770	cells/100 mL	300	cells/100mL	68.1	cells/100mL	18.2	cells/100mL	6.4	cells/100mL	0.7	cells/100 mL
	AGR 4200	880	cells/100 mL	290	cells/100mL	59.8	cells/100mL	17	cells/100mL	4.1	cells/100mL	1.6	cells/100 mL
						qPCR[a] (3 replicates)							
		Avg Ct	sd	Avg Ct	sd	Avg Ct	sd	Avg Ct	sd	Avg Ct	sd	Avg Ct	sd
lipB	AGR 4274	29.44	0.15	31.7	0.24	ND	\	ND	\	ND	\	ND	\

bglC	AGR 4200	28.55	0.13	30.57	0.53	34.31[b]	\	ND	\	ND	\	ND	\
		ddPCR (3 replicates)											
		Avg copies/100mL	sd	Avg copies/100mL	sd	Avg copies/100mL	sd	Avg copies/100mL	sd	Avg copies/100mL	sd	Avg copies/100mL	sd
lipB	AGR 4274	138.2	4.4	25.0	7.5	3.7	1.8	2.3[c]	0.9	1.6[a]	\	ND	\
bglC	AGR 4200	161.0	47.7	48.9	12.1	6.1	2.7	4.2	1.6	1.9[a]	\	ND	\

ND = Not detected; \ = not calculated; [a] 35 cycles used; [b] not detected in 2 replicates; [c] not detected in 1 replicate

3.4 Comparison between ddPCR and Colilert measurements

The recovery of target DNA during nucleic acid extraction is never complete, and failing to account for this factor can lead to misinterpretation (underestimation) of the quantitative results [52]. The DNA isolation efficiency, or ratio between the number of copies measured by ddPCR and the number of bacteria used for spiking the water samples was 18% (dilution A), and there was a consistent bias, with Colilert concentrations always measured above ddPCR concentrations (Figure 4A). However, the Bland-Altman regression analysis revealed this bias was proportional and could easily be accounted for. The agreement between the two methods was excellent with the difference D between the two methods given by $D = 1.6\,A - 0.96$ ($R^2 = 0.999$) where A is an estimation of the true concentration given by the average between the two methods (Figure 4B).

Journal Pre-proof

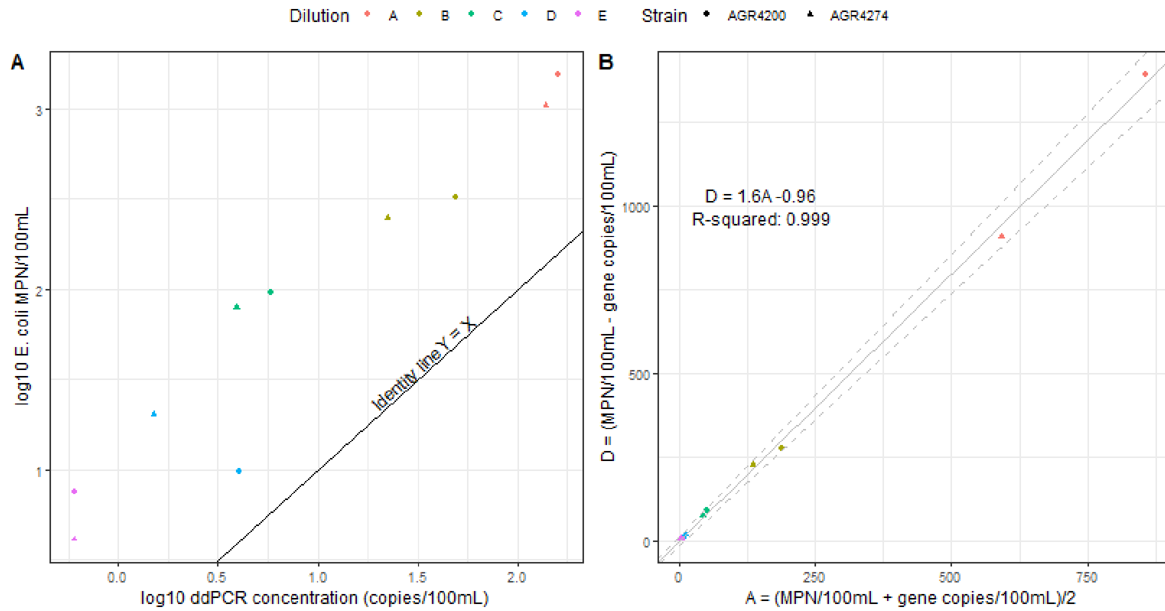


Figure 4: (A) Concentrations measured by the Most Probable Number (MPN) method (Colilert-18) and the multiplex lipB/bglC *ddPCR assay (merged concentration from triplicate measurements) of serial dilutions A to E of* E. coli *(AGR4274) and* E. marmotae *(AGR4200) spiked in water, and (B) Bland-Altman Plot of the average A between the two methods against their difference D, showing the consistency between the two methods and enabling the identification of any systematic bias or possible outliers. The middle grey line represents the regression-based proportional bias, with the dashed grey lines indicating the 95% upper and lower limits of agreement.*

Although the number of samples used at each dilution for this comparison was low, they underlined an increasing difference between the two methods that is likely related to a decrease in DNA extraction efficiency with increasing cell concentration. This issue has been noted in several microbiome studies [53, 54], but seems overlooked in quantitative molecular detection methods [55]. Future studies should aim to undertake a comprehensive comparison of conventional *E. coli* enumeration methods and the ddPCR method described herein, considering a larger sample size and this caveat, to validate the ddPCR's potential as a complementary or replacement method for current conventional methods.

An important difference between any PCR and culture-based methods lays in the detection of viable, non-viable, and free DNA, compared with the detection of viable cells only [56-58]. Agreement between the two methods will thus be impacted by whether samples tested contain a high

Journal Pre-proof

proportion of fresh fecal inputs where *E. coli* viability will better match ddPCR copies of *E. coli* DNA, as mimicked in this study by the use of fresh gDNA. This needs to be considered when interpreting and comparing results from environmental water samples where there could be a mix of fresh and non-recent faecal inputs.

3.5 Environmental water samples

Where there is no clear indication of fecal sources contributing to high *E. coli* counts in water, as in the constructed wetland example with typically 10-fold higher concentrations in the outflow compared to inflow concentrations, the visual inspection of the ddPCR 2D-amplification profile can directly provide evidence of non-*E. coli* confounding *E. coli* counts obtained by culture-based methods. The ddPCR conducted on eDNA extracted from three water samples from the Toenepi wetland confirmed that the majority of "generic *E. coli"* quantified by the Colilert method were *E. coli* (Figure 5). Whilst *E. marmotae* and *E. ruysiae* were not detected in the inflow water, they were both detected in the wetland, and *E. marmotae* was also detected at low concentration in the outlet cell (Figure 5). As previously observed [29], the concentrations of *E. coli* were higher in the middle and outlet cells samples compared to water from the inflow (Table 5).

Journal Pre-proof

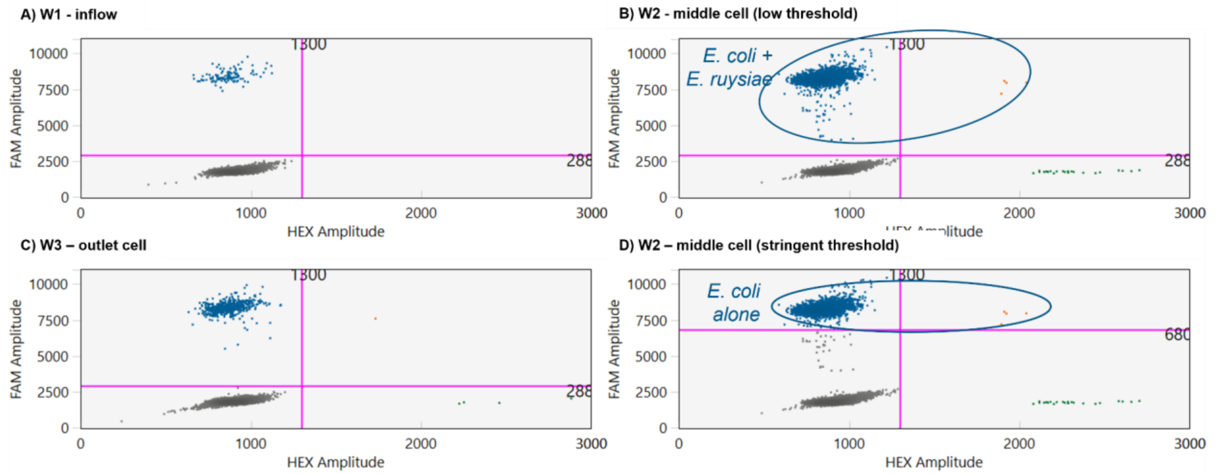


Figure 5: Visual inspection of the ddPCR 2-D Amplification graph obtained from three environmental water samples from the Toenepi constructed wetland showing A) the presence of E. coli *alone (in blue) in the wetland inflow, B)* E. marmotae *(in green),* E. coli *and* E. ruysiae *(in blue) in the wetland middle cell, and C)* E. coli *with traces of* E. marmotae *in the outlet cell. Setting B) a low threshold on the first (FAM) channel allows for quantification of* E. coli *and* E. ruysiae *together, while setting D) a stringent threshold allows the determination of* E. coli *concentration alone. Orange points represent droplets positive for both targets (*E. marmotae + E. coli ± E. ruysiae*).*

*Table 5: Generic '*Escherichia coli*' counts using a culture-based method and results of ddPCR quantification of* E. coli *and* E. ruysiae *(lipB) and* E. marmotae *(bglC) in three environmental water samples from the Toenepi constructed wetland*

	Colilert MPN/100 mL		ddPCR Copies/µl				Copies/100 mL		
Sample	*'E. coli'*	95% CI	*lipB*	95% CI	*bglC*	95% CI	*lipB*	*bglC*	*Total*
W1	86.0	[61.3 - 115.5]	3.28	[2.69-3.89]	0	[0.00-0.08]	98.4	0	98.4
W2	5475.0	[3582 - 8045]	81.3	[71.68-90.73]	0.75	[0.44-1.06]	2439.0	22.5	2461.5
W3	920.8	[620.5 - 1282.0]	16.87	[15.41-18.33]	0.16	[0.00-0.35]	506.1	4.8	510.9

The separate quantification of *E. coli* and *E. ruysiae* required the use of manual thresholding methods in the QX Manager software: first with a low threshold to obtain the overall positive droplet counts for *E. coli* and *E. ruysiae* (Figure 5B) of 2439 copies/100mL for W2, then with a more stringent threshold (Figure 5D) to obtain the concentration of *E. coli* alone of 2412 copies/100mL, to

Journal Pre-proof

finally calculate the difference to obtain *E. ruysiae* concentration, 2439 – 2412 = 27 copies/100mL. This can be subjective and may be complicated in water samples by the presence of these species at different concentrations. The use of a stringent threshold is advisable for *E. coli*, with little impact on the measured concentration when concentrations are high. However, the cluster of *E. ruysiae*-positive droplets may be difficult to differentiate from 'rain' under the cluster of *E. coli*-positive droplets when *E. ruysiae* is present at much lower concentrations than *E. coli*, and overestimation of *E. ruysiae* concentrations may occur. A similar phenomenon could also happen with the clade VI, for which no *in situ* testing was conducted in this study. This newly described clade VI [7] has, to our knowledge, not been identified in New Zealand where environmental studies have identified *E. marmotae* as the most prevalent cryptic clade, and *E. ruysiae* only rarely [13]. The rarity of *E. ruysiae* was confirmed in the present study, even in a site where it had been isolated before [13]. To improve quantification of *E. ruysiae* in areas where this species could be present at higher concentrations, the current assay could be modified into a ratio-based multiplexing assay using an *E. ruysiae*-specific probe with a different dye [59], but would likely require re-validation of the modified assay.

Despite the previous occurrence and isolation of cryptic clades in wetland samples, the ddPCR method assisted in confirming that increased export of '*E. coli*' from the wetland is not due to benign environmental *Escherichia* strains but rather more likely the extended survival, multiplication and entrainment of faecally-derived, environmentally-adapted bovine strains of *E. coli*.

3.6 Practical applications

Monitoring fecal indicator bacteria such as *E. coli* in recreational freshwaters is crucial for safeguarding public health and preserving water quality, and is the first step to detect fecal pollution, enabling timely implementation of mitigation actions [51]. Distinguishing benign, environmental *Escherichia* from fecal *E. coli* is an important step in the process. The implementation of the ddPCR —and to a certain extent the qPCR—developed in this study in water quality

assessment frameworks (i.e., elevated *E. coli* counts when there is no clear indication of fecal contamination) can be a real advantage for water quality managers, as a sensitive and rapid detection and quantification method. The culture-independent nature of this molecular method and its ability to distinguish *E. coli* from *E. marmotae* and *E. ruysiae* in a single assay provides a more rapid identification of water that is safe for recreational use than the current phenotypic methods, which can be vital for tourism and local economies that rely on freshwater activities.

Under the hypothesis that the presence of cryptic clades in water samples is associated with more pristine waters, as suggested for clade II [22], and that cryptic clades can therefore be used as indicators of water quality rather than water degradation, this tool with its potential for rare target detection could also be used to identify the presence of *E. marmotae* and *E. ruysiae* in areas where water quality is not compromised. Although not tested in this paper, the applications of this ddPCR method known for its resistance to inhibitors could also be extended to eDNA extracted from other environmental matrices (e.g., sediment, periphyton, soil), broadening the current knowledge on the ecology of environmental *Escherichia* species.

4 Conclusion

Our study demonstrated the effectiveness of qPCR and ddPCR for the simultaneous distinction and absolute quantification of *E. coli* and *E. marmotae* in water samples, and the further ability of the ddPCR assay to also distinguish *E. coli* from *E. ruysiae* and potentially other cryptic clades, in a single quantitative assay. The high sensitivity and higher tolerance to inhibitors of the ddPCR technology make it a useful tool to empower regulatory agencies and stakeholders for improved microbial water quality.

5 Acknowledgements

We are grateful for the technical support on ddPCR methodology and equipment offered by Marissa Roldan (AgResearch, Grasslands) and Sol Green (Bio-Rad). We thank Mark Mammel and David

Journal Pre-proof

Journal Pre-proof

Lacher (US Food and Drug Administration) for providing strains of *E.* ruysiae and *E. marmotae*, James P. S. Sukias (National Institute of Water and Atmospheric Research) who performed the environmental water sampling, and George Howie for access to the Toenepi constructed wetland.

Funding: This work was supported by the New Zealand Ministry of Business, Innovation and Employment Smart Idea grant number C10X1908 “Novel discriminatory tests for *E. coli* to improve water quality assessments”. The salaries of MM and RMC were supported through the AgResearch Strategic Science Investment Fund project ‘Food Integrity’.

6 Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used ChatGPT 3.5 to (i) improve readability and language in parts of this manuscript and (ii) help identify appropriate keywords. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

7 Declaration of Competing Interest

The authors declare no conflicts of interest.

8 References

1. Fewtrell, L. and J. Bartram, *Water quality : guidelines, standards and health : assessment of risk and risk management for water-related infectious diseases / edited by Lorna Fewtrell and Jamie Bartram*. 2001, Geneva: World Health Organization.
2. Wiedenmann, A., et al., *A randomized controlled trial assessing infectious disease risks from bathing in fresh recreational waters in relation to the concentration of Escherichia coli,*

Journal Pre-proof

intestinal enterococci, Clostridium perfringens, and somatic coliphages. Environ Health Perspect, 2006. **114**(2): p. 228-36.

3. Sercu, B., et al., *Cultivation-independent analysis of bacteria in IDEXX Quanti-Tray/2000 fecal indicator assays.* Appl Environ Microbiol, 2011. **77**(2): p. 627-33.
4. Manafi, M., W. Kneifel, and S. Bascomb, *Fluorogenic and chromogenic substrates used in bacterial diagnostics.* Microbiological reviews, 1991. **55**(3): p. 335-348.
5. Doyle, M.P. and J.L. Schoeni, *Survival and growth characteristics of Escherichia coli associated with hemorrhagic colitis.* Appl Environ Microbiol, 1984. **48**(4): p. 855-6.
6. Walk, S.T., *The "Cryptic" Escherichia.* EcoSal Plus, 2015. **6**(2).
7. Gangiredla, J., et al., *Draft genome sequences of Escherichia albertii, Escherichia fergusonii, and strains belonging to six cryptic lineages of Escherichia spp.* Genome Announc, 2018. **6**(18): p. e00271-18.
8. Zhou, Z., et al., *The EnteroBase user's guide, with case studies on Salmonella transmissions, Yersinia pestis phylogeny, and Escherichia core genomic diversity.* Genome Res, 2020. **30**(1): p. 138-152.
9. Gilroy, R., et al., *Extensive microbial diversity within the chicken gut microbiome revealed by metagenomics and culture.* PeerJ, 2021. **9**: p. e10941.
10. Liu, S., et al., *Genomic and molecular characterisation of Escherichia marmotae from wild rodents in Qinghai-Tibet plateau as a potential pathogen.* Scientific Reports, 2019. **9**(1).
11. van der Putten, B.C.L., et al., *Escherichia ruysiae sp. nov., a novel Gram-stain-negative bacterium, isolated from a faecal sample of an international traveller.* Int J Syst Evol Microbiol, 2021. **71**(2).
12. Sivertsen, A., et al., *Escherichia marmotae-a Human Pathogen Easily Misidentified as Escherichia coli.* Microbiol Spectr, 2022. **10**(2): p. e0203521.

Journal Pre-proof

13. Cookson, A.L., et al., *Whole-genome sequencing and virulome analysis of Escherichia coli isolated from New Zealand environments of contrasting observed land use.* Appl Environ Microbiol, 2022. **88**(9): p. e00277-22.
14. Moinet, M., et al., *High-resolution genomic analysis to investigate the impact of the invasive brushtail possum (Trichosurus vulpecula) and other wildlife on microbial water quality assessments.* PLOS ONE, 2024. **19**(1): p. e0295529.
15. Biggs, P.J., et al., *Draft genome sequences of Escherichia spp. isolates from New Zealand environmental sources.* Microbiology Resource Announcements, 2024. p. e0100723
16. Jang, J., et al., *Environmental Escherichia coli: ecology and public health implications-a review.* J Appl Microbiol, 2017. **123**(3): p. 570-581.
17. NZ Ministry for the Environment, *National Policy Statement for Freshwater Management 2020*. 2022, Minister for the Environment.
18. Walk, S.T., et al., *Genetic diversity and population structure of Escherichia coli isolated from freshwater beaches.* Environ Microbiol, 2007. **9**(9): p. 2274-88.
19. Blyton, M.D.J., et al., *Genetic structure and antimicrobial resistance of Escherichia coli and cryptic clades in birds with diverse human associations.* Appl Environ Microbiol, 2015. **81**(15): p. 5123-5133.
20. Vignaroli, C., et al., *Adhesion of marine cryptic Escherichia isolates to human intestinal epithelial cells.* The ISME Journal, 2015. **9**(2): p. 508-515.
21. Quero, G.M., et al., *Understanding the association of Escherichia coli with diverse macroalgae in the lagoon of Venice.* Sci Rep, 2015. **5**: p. 10969.
22. Koh, X.P., et al., *Genetic and Ecological Diversity of Escherichia coli and Cryptic Escherichia Clades in Subtropical Aquatic Environments.* Frontiers in Microbiology, 2022. **13**.
23. Walk, S.T., et al., *Cryptic lineages of the genus Escherichia.* Appl Environ Microbiol, 2009. **75**(20): p. 6534-6544.

Journal Pre-proof

24. Clermont, O., et al., *Characterization of the cryptic Escherichia lineages: rapid identification and prevalence.* Environ Microbiol, 2011. **13**(9): p. 2468-77.
25. Sabarly, V., et al., *The decoupling between genetic structure and metabolic phenotypes in Escherichia coli leads to continuous phenotypic diversity.* Journal of Evolutionary Biology, 2011. **24**(7): p. 1559-1571.
26. Clermont, O., et al., *The Clermont Escherichia coli phylo-typing method revisited: improvement of specificity and detection of new phylo-groups.* Environmental Microbiology Reports, 2013. **5**(1): p. 58-65.
27. Mire, M., et al., *Escherichia cryptic clade II through clade VIII: Rapid detection and prevalence in feces and surface water.* Sci Total Environ, 2022. **848**: p. 157741.
28. Deng, D., et al., *Differentiating enteric Escherichia coli from environmental bacteria through the putative glucosyltransferase gene (ycjM).* Water Res, 2014. **61**: p. 224-31.
29. Stott, R., et al., *Net export of E. coli from a constructed wetland treating agricultural drainage.* Ecol Eng, 2023. **194**: p. 107023.
30. Zhi, S., et al., *Evidence for the evolution, clonal expansion and global dissemination of water treatment-resistant naturalized strains of Escherichia coli in wastewater.* Water Res, 2019. **156**: p. 208-222.
31. Zhi, S., et al., *Evidence of naturalized stress-tolerant strains of Escherichia coli in municipal wastewater treatment plants.* Applied and Environmental Microbiology, 2016. **82**(18): p. 5505-5518.
32. Hindson, C.M., et al., *Absolute quantification by droplet digital PCR versus analog real-time PCR.* Nature Methods, 2013. **10**(10): p. 1003-1005.
33. Dingle, T.C., et al., *Tolerance of Droplet-Digital PCR vs Real-Time Quantitative PCR to Inhibitory Substances.* Clinical Chemistry, 2013. **59**(11): p. 1670-1672.

Journal Pre-proof

34. Bian, X., et al., *A microfluidic droplet digital PCR for simultaneous detection of pathogenic Escherichia coli O157 and Listeria monocytogenes.* Biosensors and Bioelectronics, 2015. **74**: p. 770-777.

35. Capobianco, J.A., et al., *Detection of Shiga toxin-producing Escherichia coli (STEC) in beef products using droplet digital PCR.* International journal of food microbiology, 2020. **319**: p. 108499.

36. He, L., D.J. Simpson, and M.G. Gänzle, *Detection of enterohaemorrhagic Escherichia coli in food by droplet digital PCR to detect simultaneous virulence factors in a single genome.* Food microbiology, 2020. **90**: p. 103466.

37. Verhaegen, B., et al., *Comparison of droplet digital PCR and qPCR for the quantification of shiga toxin-producing Escherichia coli in bovine feces.* Toxins, 2016. **8**(5): p. 157.

38. Suzuki, Y., et al., *Simultaneous detection of various pathogenic Escherichia coli in water by sequencing multiplex PCR amplicons.* Environmental Monitoring and Assessment, 2023. **195**(2): p. 264.

39. Mancusi, A., et al., *Droplet Digital PCR (ddPCR) Analysis for Detecting Shiga-Toxin-Producing Escherichia coli (STEC).* Applied Sciences (Switzerland), 2022. **12**(7).

40. Collis, R., *Metagenomic analysis and culture-based methods to examine the prevalence and distribution of antimicrobial resistance on two New Zealand dairy farms : a thesis presented in partial fulfilment of the requirements for the degree of Doctor of Philosophy in Veterinary Science at Massey University, Palmerston North, New Zealand*. 2022, Massey University.

41. Wilhelm, S., *M9 Minimal Media.* protocols.io, 2017.

42. Miles, A.A., S.S. Misra, and J.O. Irwin, *The estimation of the bactericidal power of the blood.* Journal of Hygiene, 1938. **38**(6): p. 732-749.

43. Beghain, J., et al., *ClermonTyping: an easy-to-use and accurate in silico method for Escherichia genus strain phylotyping.* Microb Genom, 2018. **4**(7).

Journal Pre-proof

44. Klymus, K.E., et al., *Reporting the limits of detection and quantification for environmental DNA assays.* Environmental DNA, 2020. **2**(3): p. 271-282.

45. Bustin, S.A., et al., *The MIQE Guidelines: Minimum Information for Publication of Quantitative Real-Time PCR Experiments*. 2009, Oxford University Press.

46. Bland, J.M. and D.G. Altman, *Measuring agreement in method comparison studies.* Statistical Methods in Medical Research, 1999. **8**(2): p. 135-160.

47. Wickham, H., *ggplot2 Elegant graphics for data analysis*. Second Edition ed. Use R! 2016: Springer Cham.

48. Pédron, J. and F. Van Gijsegem, *Diversity in the Bacterial Genus Dickeya Grouping Plant Pathogens and Waterways Isolates.* OBM Genetics, 2019. **03**(04): p. 098.

49. Hugouvieux-Cotte-Pattat, N., J. Pédron, and F. Van Gijsegem, *Insight into biodiversity of the recently rearranged genus Dickeya.* Frontiers in Plant Science, 2023. **14**.

50. Parkinson, N., et al., *Dickeya aquatica sp. nov., isolated from waterways.* Int J Syst Evol Microbiol, 2014. **64**(Pt 7): p. 2264-2266.

51. World Health Organisation, *Guidelines for safe recreational water environments Volume 1: Coastal and freshwaters*. 2021, World Health Organisation: Geneva.

52. Kralik, P. and M. Ricchi, *A Basic Guide to Real Time PCR in Microbial Diagnostics: Definitions, Parameters, and Everything.* Frontiers in Microbiology, 2017. **8**.

53. Davis, A., et al., *Improved yield and accuracy for DNA extraction in microbiome studies with variation in microbial biomass.* BioTechniques, 2019. **66**(6): p. 285-289.

54. Guo, F. and T. Zhang, *Biases during DNA extraction of activated sludge samples revealed by high throughput sequencing.* Appl Microbiol Biotechnol, 2013. **97**(10): p. 4607-16.

55. Luk, A.W., S. Beckmann, and M. Manefield, *Dependency of DNA extraction efficiency on cell concentration confounds molecular quantification of microorganisms in groundwater.* FEMS Microbiology Ecology, 2018. **94**(10).

Journal Pre-proof

56. Bae, S. and S. Wuertz, *Discrimination of viable and dead fecal Bacteroidales bacteria by quantitative PCR with propidium monoazide.* Appl Environ Microbiol, 2009. **75**(9): p. 2940-4.

57. Oliver, J.D., *Recent findings on the viable but nonculturable state in pathogenic bacteria.* FEMS Microbiol Rev, 2010. **34**(4): p. 415-25.

58. Weigel, K.M., et al., *Molecular Viability Testing of UV-Inactivated Bacteria.* Appl Environ Microbiol, 2017. **83**(10).

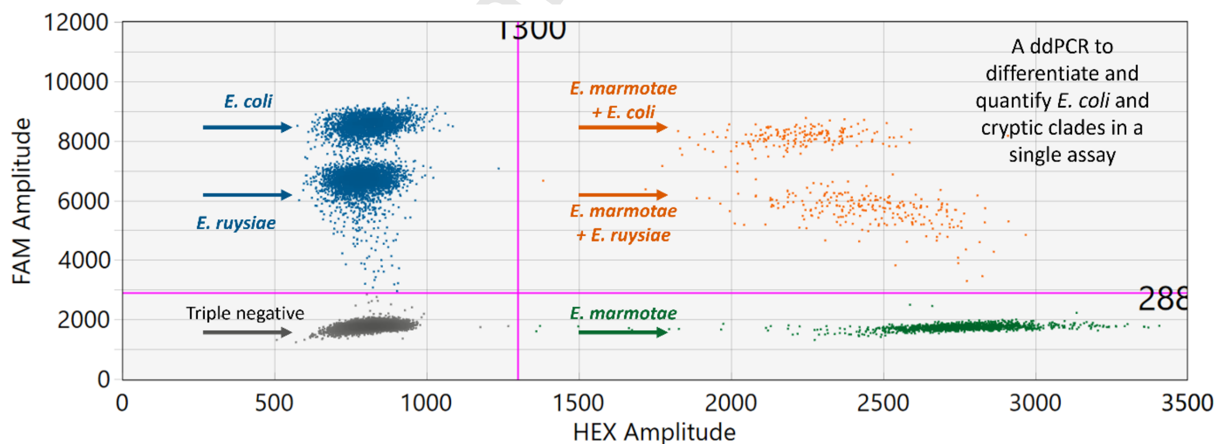
59. Whale, A.S., J.F. Huggett, and S. Tzonev, *Fundamentals of multiplexing with digital PCR.* Biomol Detect Quantif, 2016. **10**: p. 15-23.

Journal Pre-proof

Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Graphical abstract



Highlights

- Novel droplet digital PCR assay for environmental DNA applications
- Ability to differentiate and quantify *E. coli* and cryptic clades in a single run
- Specific quantitation of fecal *E. coli* in water with high sensitivity and precision

- Reduced turnaround time for in-depth microbial water quality assessments

Journal Pre-proof