The Journal of Molecular Diagnostics, Vol. 27, No. 6, June 2025

ELSEVIER

the Journal of Molecular Diagnostics
jmdjournal.org

High-Throughput Multiplex Detection of Antibiotic-Resistant Genes and Virulence Factors in *Escherichia coli* Using Digital Multiplex Ligation Assay

Sheena Conforti,*†‡ Pablo Rossi Orts,§ Manu Tamminen,¶ and Timothy R. Julian*‖**

From Eawag, Swiss Federal Institute of Aquatic Science and Technology, Duebendorf, Switzerland; the Department of Biosystems Science and Engineering,† ETH Zurich, Basel, Switzerland; the Swiss Institute of Bioinformatics,‡ Lausanne, Switzerland; the Department of Environmental Systems Science,§ ETH Zurich, Zurich, Switzerland; the Department of Biology,¶ University of Turku, Turku, Finland; the Swiss Tropical and Public Health Institute,‖ Allschwil, Switzerland; and the University of Basel,** Basel, Switzerland*

Accepted for publication March 5, 2025.

Address correspondence to Timothy R. Julian, Ph.D., Department of Environmental Microbiology, Eawag, Swiss Federal Institute of Aquatic Science and Technology, Ueberlandstr. 133, 8600 Duebendorf, Switzerland. E-mail: tim.julian@eawag.ch.

Escherichia coli causes >400,000 annual deaths in children aged <5 years worldwide, with morbidity and mortality exacerbated by antimicrobial-resistant strains. A high-throughput multiplexing assay called digital multiplex ligation assay (dMLA) was developed to detect simultaneously 43 priority genes in *E. coli* related to the following: antibiotic resistance ($n = 19$), virulence factors ($n = 16$), and phylogroup markers ($n = 6$) with controls (*uidA*, *gapdh*). Genes are detected via PCR amplification of adjacent probe pairs that ligate in the presence of target gene-specific DNA, followed by sequencing of amplicons on short-read sequencers. The assay was tested in technical replicates on 63 synthetic DNA controls, and applied to 58 *E. coli*, 2 *Staphylococcus aureus*, 2 *Klebsiella pneumoniae*, 1 *Klebsiella oxytoca*, 1 *Vibrio cholera*, 1 *Pseudomonas lurida*, and 1 *Salmonella enterica* isolates in duplicate. Whole-genome sequencing was used to assess specificity and sensitivity. dMLA showed 100% sensitivity and >99.9% specificity and balanced accuracy on synthetic DNA. Balanced accuracy, calculated as the average of sensitivity and specificity, accounts for imbalanced data sets where negative outcomes are significantly more prevalent than positive ones. dMLA achieved a balanced accuracy of 90% for bacterial isolates. The results underline dMLA's effectiveness in high-throughput characterization of *E. coli* libraries for antimicrobial resistance genes and virulence factors, leveraging sequencing for massively parallel multiplexing of gene regions on multiple samples simultaneously, and are extendable to targets beyond *E. coli*. *(J Mol Diagn 2025, 27: 511—524; https://doi.org/10.1016/j.jmoldx.2025.03.003)*

The vast majority of *Escherichia coli* are harmless commensals living within the human gut, but a critical fraction are pathogenic.[1] Pathogenic *E. coli* are among the leading cause of moderate-to-severe diarrhea in low-resource settings,[2] and diarrheal diseases are the major cause of childhood mortality, with yearly estimated 444,000 deaths in children aged <5 years (*https://data.unicef.org/topic/child-health/diarrhoeal-disease*, last accessed April 14, 2024). Intestinal pathogenic *E. coli* strains are classified into six pathotypes: enteroaggregative, enterohemorrhagic, enteroinvasive, enteropathogenic, enterotoxigenic, and diffusely adherent *E. coli*.[3] Pathotypes are distinguished by virulence factors (VFs), which are small molecules and proteins that enable *E. coli* to cause disease by affecting a wide range of cellular processes.[1]

The emergence of antimicrobial resistance (AMR) in pathogenic *E. coli* strains presents an additional challenge, complicating treatment options and leading to increased morbidity, mortality, and health care expenditures.[4] Low- and middle-income countries are the most burdened with both diarrheal diseases and antimicrobial resistance carriage, but they often have insufficient capacity to establish and/or

Supported in whole or in part by the Swiss National Science Foundation grant 192763 (T.R.J.).

https://doi.org/10.1016/j.jmoldx.2025.03.003

maintain surveillance systems,[5,6] leaving much of the AMR burden in these settings unquantified.[7]

Surveillance of AMR and pathogenicity is conducted through global and regional networks, including clinical and environmental assessments.[8] The World Health Organization Global Antimicrobial Resistance and Use Surveillance System is a key example, promoting standardized global collection and sharing of AMR data.[9] Beyond clinical sources, the Global Antimicrobial Resistance and Use Surveillance System has implemented a One Health approach with the Tricycle Surveillance Project, which integrates data from humans, animals, and environmental sources.[10] Additional networks include the CDC Gonococcal Isolate Surveillance Program in the United States (*https://stacks.cdc.gov/view/cdc/40081*, last accessed March 13, 2025), the European Antimicrobial Resistance Surveillance Network in Europe,[11] and the Central Asian and European Surveillance of Antimicrobial Resistance network,[12] all aligning with the World Health Organization initiatives. The academic and private sectors also significantly enhance AMR surveillance.[13]

AMR surveillance incorporates both phenotypic and genotypic approaches. Predominantly, the above-mentioned networks generate culture-based AMR data through phenotypic assays, which evaluate antimicrobial effectiveness against bacteria, assisting in therapeutic guidance and resistance mapping.[14] Nevertheless, culture-based methods are limited in the range of antibiotics they can test, are time-consuming and resource intensive, and cannot distinguish between pathogenic and nonpathogenic *E. coli* strains.[15] Therefore, comprehensive *E. coli* characterization requires methods beyond culture-based protocols. Genotypic techniques, such as quantitative real-time PCR, digital PCR, multiplex PCR, and sequencing, offer rapid detection of resistance genes and insights into resistance mechanisms and VFs not identifiable by phenotypic tests. However, despite their precision, genotypic methods are constrained by their capacity to identify a limited number of targeted DNA regions simultaneously, presenting a significant challenge for large-scale surveillance.[16–18] Metagenomics offers a comprehensive overview of microbial communities, including nonculturable bacteria, but its high costs, limited sensitivity, and data complexity limit its practicality for routine monitoring.[19,20]

The digital multiplex ligation assay (dMLA) is a low-cost, high-throughput molecular technique initially developed by Tamminen et al,[21] in 2020, to detect β-lactamase genes. This assay offers a cost-effective and efficient approach for screening large isolate collections, such as those stored in archives, or obtained within the scope of monitoring (eg, nationally representative household water quality testing, such as the Multiple Indicator Cluster Survey).[22] By using *in silico* probe design, ligation-based detection, and next-generation sequencing, the multiplexing ability of dMLA significantly reduces costs and enhances throughput. This enables comprehensive surveillance of potentially hundreds of genes across hundreds of bacterial isolates in a single sequencing effort.

In this study, the application of dMLA was extended beyond β-lactamase genes to include 43 genes related to antibiotic resistance, VFs, and *E. coli*–specific genetic markers. This expansion aims at the comprehensive characterization of *E. coli* isolates through identification of pathotypes, resistance across six antibiotic classes, phylogroup classification, and species-level confirmation.

Materials and Methods

Selection of Target Genes and Retrieval of Reference Sequences

Antibiotic-Resistant Genes

Nineteen[19] antibiotic-resistant genes were selected on the basis of their prevalence in human-associated environments, their potential for horizontal transfer among bacteria, their presence in ESKAPE (*Enterococcus faecium*, *Staphylococcus aureus*, *Klebsiella pneumoniae*, *Acinetobacter baumannii*, *Pseudomonas aeruginosa*, and *Enterobacter* spp.) pathogens,[23] or their clinical threat due to rapid emergence and diffusion, as highlighted in multiple studies[24–27] (Table 1). Reference sequences were obtained from the Comprehensive Antibiotic Resistance Database (*https://card.mcmaster.ca*, last accessed September 29, 2022), focusing only on homologs related to *E. coli*, the target organism. The selection process for downloading specific homologs from the Comprehensive Antibiotic Resistance Database involved verifying the presence of *E. coli* in either resistomes with perfect matches or resistomes with sequence variants. If *E. coli* was not identified in either category, only those homologs were included that demonstrated a prevalence of >1.5% in whole-genome shotgun assemblies within the National Center for Biotechnology Information (NCBI) database (*https://www.ncbi.nlm.nih.gov*, last accessed February 17, 2021). The comprehensive list of chosen homologs is available in Supplemental Table S1.

Virulence Factors

Sixteen[16] VFs associated with five *E. coli* pathotypes—enteroaggregative *E. coli*, enterohemorrhagic *E. coli*, enteroinvasive *E. coli*, enteropathogenic *E. coli*, and enterotoxigenic *E. coli*—were selected (Table 1). Reference sequences for these VFs were retrieved from the Virulence Factors Database (*http://www.mgc.ac.cn/VFs*, last accessed January 23, 2021).[28] The full data set of verified and predicted VFs was downloaded from the Virulence Factors Database. After filtering of the data set to include only *E. coli* sequences, the VFs of interest were identified by the sequence titles, and the corresponding available references for each VF were downloaded. Between 1 and 15 reference sequences were obtained for each VF, with accession numbers detailed in Supplemental Table S2. These

Table 1 Target Genes for Antibiotic Resistance, Virulence, and Phylogroup Identification in *Escherichia coli* Using Digital Multiplex Ligation Assay

Resistance	Antibiotic class	Antibiotic-resistant gene
	Aminoglycosides	*aadA, aph(3′)-I, aac(6′)-I, aac(3)-II, aac(3)-VI, ant(2″)-Ia, aph(6)-I*
	Colistin	*mcr*
	Quinolones	*qnrB*, *qnrS*
	Macrolides	*ermB*, *mphA*
	Sulfonamides/diaminopyrimidines	*dfrA*, *sul1*, *sul2*, *sul3*
	Tetracycline	*tetA*, *tetB*, *tetM*
Pathogenicity	Pathotype	Virulence factor
	EAEC	*astA*, *aggR*, *aatA*, *aafA*, *aap*, *aaiC*
	EPEC	*bfpA*, *bfpF*
	EIEC	*invE*, *ipaH*
	EPEC/EHEC	*eae*
	EHEC	*stx1*, *stx2*
	ETEC	*eltA, eltB, estIa*
Phylogroup	Genetic marker	
	chuA, *yjaA*, *trpA*, *arpA*, TspE4.C2	
E. coli positive control	*uidA*	
E. coli negative control	*gapdh*	
Resistance marker	*intI1*	

This table categorizes the selected genes into three main sections: antibiotic resistance genes across six antibiotic classes, virulence factors for five *E. coli* pathotypes, and genetic markers for *E. coli* phylogroup identification, including positive and negative controls for *E. coli* detection and a resistance marker.
EAEC, enteroaggregative *E. coli*; EHEC, enterohemorrhagic *E. coli*; EIEC, enteroinvasive *E. coli*; EPEC, enteropathogenic *E. coli*; ETEC, enterotoxigenic *E. coli*.

sequences were then blasted against the NCBI nucleotide database to identify closely related *E. coli* sequences. The Basic Local Alignment Search Tool (BLAST; *https://blast.ncbi.nlm.nih.gov/Blast.cgi*) search parameters were set to yield a maximum of 5000 target sequences, maintaining other parameters as defaults, including an expected threshold of 0.05, a word size of 28, and no maximum matches in a query range. The resulting sequences for each VF were compiled into a unified data set, duplicates were removed, and ultimately, 16 data sets with unique sequence (one for each VF) were prepared for probe-pair development.

Phylogroup Genetic Markers

Five phylogroup genetic markers (*trpA*, *chuA*, *arpA*, *TspE4.C2*, and *yjaA*) were selected in accordance with the Clermont et al,[29] 2013, protocol to classify *E. coli* into corresponding phylogroups. Unique reference sequences for each of these genes were obtained from the NCBI nucleotide database (*https://www.ncbi.nlm.nih.gov*, last accessed February 17, 2021) (Supplemental Table S3). These sequences were subjected to BLAST analysis against the NCBI nucleotide database to identify closely related sequences, which were subsequently downloaded. The BLAST search parameters were set to yield a maximum of 1000 target sequences, maintaining other parameters as defaults, including an expected threshold of 0.05, a word size of 28, and no maximum matches in a query range. After running the BLAST search, the results were filtered for the organism *Escherichia coli*, and all corresponding sequences were selected. The sequences corresponding to each phylogroup genetic marker were downloaded, resulting in a total of five data sets. These data sets were then filtered to retain only unique sequences to reduce processing time and to limit redundant results. The filtered data sets, characterized by unique sequences only, were used for *in silico* development of probe-pairs.

Controls and *intI1* Marker

The dMLA was designed to also include *intI1*, which encodes for integrase and is often associated with antibiotic-resistant genes and horizontal gene transfer,[18,30] *uidA* (used for *E. coli* species confirmation as a positive control), and *gadph* (used as a negative control because it is absent in *E. coli*, but present in *Klebsiella pneumoniae*). Unique reference sequences for *uidA*, *gadph*, and *intI1* were obtained from the NCBI nucleotide database, and the same BLAST analysis procedures were followed as described in *Phylogroup Genetic Markers* (*https://www.ncbi.nlm.nih.gov*, last accessed February 17, 2021) (Supplemental Table S3). After filtering the results specific to each organism, the corresponding sequences were downloaded. For the sequences of the *gadph* gene, the results were filtered for the organism *K. pneumoniae*. These data sets, now including *intI1*, *uidA*, and *gadph*, were also filtered to include only unique sequences, forming three additional data sets used to develop the probe-pairs targeting the corresponding genes.

In Silico Probe Design

Probe pairs for each of the 43 genes were designed using the prider package in R statistical software version 4.2.1 (*http://pride-r.github.io/prideR*).[31] Each data set of sequences, downloaded and generated as previously described, was analysed with *prider*. The parameters used in *prider* were as follows: cumulative coverage decimals set to one, a minimum primer group size of three, minimum sequence group size of one, primer length at 40 nucleotides, and GChalves option activated. For each gene, one to five sequence clusters were identified to ensure maximum coverage of the reference sequence while minimizing sequence overlap. From these clusters comprising multiple 40-bp probes that fit the specified criteria, up to three probes per cluster were selected for experimental validation. Each probe was then divided to generate two half-probes of 20 bp each. The 5' end of the left half-probe was modified by adding the sequence TGGGCCCAATTTTCCGTGACAATTAATT, which incorporates the primer binding site (underlined). Similarly, the 3' end of the right half-probe was extended with the sequence NNNNNNNNNNGAATGAGTGTGCGTGCACTC, which includes the reverse primer binding site (underlined), and NNNNNNNNNN serves as a molecular barcode to quantify unique probe molecules in a DNA sample and adjust for PCR amplification bias. Furthermore, a reverse complementary sequence of the original 40-bp probe was used as a positive control in the probe testing.

Probe Testing and Validation of dMLA

Single Plex

After designing the probes *in silico*, left and right half-probes, along with their reverse complements serving as a positive control, were synthesized (Microsynth AG, Balgach, Switzerland), and validated experimentally. Probe pairs were tested in single plex through a ligation reaction followed by a PCR, as previously described.[21] Ligation was conducted using a Biometra T3000 thermocycler (Analytik Jena, Jena, Germany), programmed to 94°C for 10 minutes, followed by 60°C for 90 minutes, containing a total reaction volume of 25 μL, including the following: 1 μL of AmpLigase (5 units; Epicentre and Lucigen; LGC Biosearch Technologies, Middleton, WI), 2.5 μL of AmpLigase Buffer (10×; Epicentre and Lucigen), 1 μL of left probe (1 μmol/L), 1 μL of right probe (1 μmol/L), and template. The template consisted of any of the following: i) 1 μL of positive control DNA template (the reverse complement sequence of both probes, 10 pmol/L), ii) 1 μL of DNase/RNase-free water as a ligation no template control (NTC), or iii) 1 μL of DNA extract from bacteria known to contain the target gene. The subsequent PCR (25 μL total volume) included 0.25 μL of Phusion High-Fidelity DNA Polymerase (2 units; New England BioLabs, Ipswich, MA), 5 μL of Phusion buffer (1×; BioLabs), 0.5 μL of dNTPs (10 nmol/L), 1 μL of forward primer (10 μmol/L; 5′-TCTTTTCGCAGGCTGGAGCCCAGGTCTTCCTATA-TAAGACTGGGCCCAATTTTCCGTGAC-3′), 1 μL of reverse primer (10 μmol/L; 5′-GCTGAACCGCTC TTCCGGTATCATCTTCTCCAAATGGGTCATGATC-3′), and either 1 μL of the ligation product or 1 μL of DNase/RNase-free water as PCR negative control. Thermocycling was set to 98°C for 30 seconds, then 35 cycles of 98°C for 5 seconds, 70°C for 1 second, and 72°C for 3 seconds, with a final extension at 72°C for 5 minutes. PCR products (10 μL) were analyzed on a 2.5% agarose gel in TAE Buffer using 0.01% SYBR SAFE (Thermo Fisher Scientific, Schlieren, Switzerland) on a Biorad PowerPac 300 Gel Electrophoresis and multiSUB Choice Wide Midi Horizontal Electrophoresis System, alongside 3 μL of a 100-bp standard ladder (Promega, Madison, WI). Probe-pairs were classified as unsuccessful if no band appeared at 140 bp for the positive control or if such a band was observed in either the NTC or in the PCR negative control (Supplemental Figure S1). If one probe-pair was classified as unsuccessful, another probe-pair from the same sequence cluster was selected for experimental validation. The 63 successful probe-pairs targeting 43 genes are listed in Supplemental Table S4.

Multiplex on Synthetic DNA (Positive Controls)

Following successful single-plex performance, the 63 successful probe-pairs were selected for multiplex testing against their specific reverse cDNA templates, which served as positive controls, alongside eight NTCs to assess assay specificity. Multiplex testing was performed on two technical replicates. The procedure involved mixing 1 μL of each half probe-pair from a 100 μmol/L stock to generate a probe mix, which was then used in ligation reactions. Ligation reactions comprised 1 μL of AmpLigase (5 units; LGC Biosearch Technologies), 2.5 μL of AmpLigase Buffer (10×; LGC Biosearch Technologies), 1 μL of the probe mix to achieve a final concentration of 1 μmol/L for each half-probe, and 1 μL of an artificial DNA template (10 pmol/L) or DNase/RNase-free water for the NTCs, in a total reaction volume of 25 μL. Following ligation, PCR was conducted on the products, incorporating one PCR-negative control where DNase/RNase-free water replaced the ligation product to evaluate PCR specificity. Unique barcoded forward primers for each sample (positive control DNA templates or NTCs) and a uniform reverse primer were used in the PCR (Supplemental Table S5). Subsequently, 15 μL of each uniquely barcoded PCR product was pooled and purified using the Wizard Genomic DNA Purification Kit, followed by sequencing of the pooled PCR products on an Illumina NovaSeq platform (Eurofins Genomics, GmbH, Ebersberg, Germany) to generate 150-bp paired-end reads.

Application of dMLA to Bacterial Isolates

The dMLA was applied to 66 bacterial isolates, each in duplicates, including the following: 58 *E. coli*, 2 *Staphylococcus aureus*, 2 *K. pneumoniae*, 1 *Klebsiella oxytoca*, 1 *Vibrio cholera*, 1 *Pseudomonas lurida*, and 1 *Salmonella enterica*. Whole-genome sequences of the 66 bacterial isolates were available for subsequent evaluation of assay performance (Supplemental Table S6). Three different reactions were performed. The first reaction included 30 *E. coli* isolates, tested in duplicate, along with one positive control (probe-pair uidA160) in duplicate, eight negative controls where DNase/RNase-free water was used, and one PCR-negative control with DNase/RNase-free water replacing the ligation product to assess PCR specificity. The second reaction involved 28 *E. coli* isolates, also tested in duplicate, with two positive controls (probe-pairs aaiC47 and aggR11) in duplicate, eight negative controls using DNase/RNase-free water, and two PCR-negative controls to evaluate PCR specificity. The third reaction included the eight non−*E. coli* isolates, tested in duplicate, two positive controls (probe-pairs aaiC47 and aggR11) in duplicate, 48 negative controls using DNase/RNase-free water, and two PCR-negative controls to evaluate PCR specificity. DNA was extracted from the bacterial isolates using the Qiagen (Hilden, Germany) Blood and Tissue kit, according to the manufacturer's instructions, starting from 1 mL of enriched liquid culture in Luria Broth (AppliChem GmbH, Darmstadt, Germany). The quality and concentration of the extracted DNA were measured using a NanoDrop spectrophotometer (Thermo Fisher Scientific), and DNA concentration was not normalized relative to volume before using the DNA in the assay (Supplemental Table S7). The purification of PCR products and subsequent sequencing followed the same protocol as described for the synthetic controls, using the Wizard Genomic DNA Purification Kit and Illumina NovaSeq platform to generate 150-bp paired-end reads.

Bacterial Sample Selection

Escherichia coli isolates were selected on the basis of the availability of an isolate in the laboratory with complementary whole-genome sequencing for confirmation of presence/absence of gene targets, ensuring geographic and source-based variability. Thirty isolates were collected from cattle, soil, humans, and chickens in rural villages of Bangladesh between February and April 2016,[32] and 28 isolates were collected from Swiss wastewater in October 2023. Non−*E. coli* isolates were included to assess the specificity of the assay for *E. coli*, particularly targeting the *uidA* gene to confirm species identity and the *gapdh* gene as a negative control, expected to be present only in *Klebsiella* species. The two *S. aureus* strains were collected from Swiss wastewater in December 2023 and January 2024. The *K. pneumoniae* isolates were provided by Patrice Nordmann's laboratory (University of Fribourg, Fribourg, Switzerland). The *K. oxytoca*, *V. cholera*, and *S. enterica* isolates were provided by Microbial Systems Ecology Group (Eawag, Duebendorf, Switzerland), whereas the *P. lurida* isolate was provided by Frederik Hammes' laboratory (Eawag, Duebendorf, Switzerland).

Sequence Data Analysis

Next-generation sequencing reads in FASTQ format were merged using VSEARCH version 2.19.0.[33] The merged reads were processed into unique molecular barcodes counts or unique molecular identifier (UMI) counts using an ad hoc Python version 3.10.10 script (*https://www.python.org*) by matching the targets according to the forward primer barcode and the target probe-pair. Data were visualized using the package ggplot2 version 3.4.2 in R version 4.1.1 via RStudio version 2024.04.2 (*https://posit.co/download/rstudio-desktop*).

After sequencing, UMI counts for target molecules were observed in samples where they were not expected, indicating the presence of false positives (FPs). To minimize these FPs, a target threshold was set. All calculations were performed in R version 4.1.1, with data manipulation using the *dplyr* package version 1.1.1 and threshold calculation performed using base R functions. The threshold was defined on the basis of the γ distribution of FP UMI counts for each target gene. Shape and rate parameters of the γ distribution were estimated using the method of moments, derived from the mean and variance of UMI counts identified as FP. These UMI counts were determined in instances where reads corresponding to a target gene were found either exclusively in the presence of a positive template of a different target gene or in samples designated as negative (NTCs), thus considered FPs. Estimates were used to calculate the 99.9% quantile threshold for the γ distribution associated with each target. Subtraction of this threshold from the original count of each target in each sample was performed, refining the detection to include only the upper 0.1% of positive reads. Any resulting negative values from this subtraction were adjusted to 0. The analyses were conducted separately for the two technical replicates and the three reactions performed on bacterial isolates. After completing the individual analyses, the results of the three reactions on bacterial isolates were combined to evaluate the performance of the assay and presented in a single figure for comprehensive visualization and comparison.

Comparison of dMLA and Whole-Genome Sequencing Results of *E. coli* Isolates

Probe-pairs were blasted against the assembled scaffolds of the whole-genome sequences of the *E. coli* isolates using BLAST + version 2.15.0. The sequences of *E. coli* from set 1 were retrieved from a previous study.[32] The Illumina sequencing of the isolates of *E. coli* from set 2 was performed using the LITE pipeline on a NextSeq 1000 P2 flow cell (300-bp paired end) at the Earlham Institute (Norwich, UK). Sequences are available on DBJ/ENA/GenBank under the accession numbers VNWZ00000000 to

VNZG00000000[32] and at the submission SUB14702901 and SUB14920117 under the BioProject PR JNA1155916 (*https://www.ncbi.nlm.nih.gov/bioproject/527972*, last accessed April 17, 2025) (Supplemental Table S6). A probe-pair was considered present if it matched a region within the assembled genome with a percentage similarity of ≥95%.

Assay Performance Evaluation

The number of FPs, false negatives (FNs), true positives (TPs), and true negatives (TNs) in each reaction was calculated in R version 4.1.1. The number expected positives (EPs) and expected negatives (ENs) was determined on the basis of the presence of the genomic region targeted by the probe-pairs in the whole genome sequences of the *E. coli* isolates. Five distinct statistical metrics were used to assess the predictive performance of the molecular targeted assay: sensitivity, specificity, precision, F1 score, and balanced accuracy. All calculations were performed in R version 4.1.1.

Sensitivity measures the proportion of TPs, and it was calculated to quantify dMLA's ability to determine TPs among the isolates.[34] Sensitivity was calculated as the number of TPs divided by the sum of TP and FNs.[34] Precision is defined as the proportion of predicted positives that are TPs or predicted negatives that are TNs. It indicates how often the assay correctly predicts either a positive or negative result.[34] Precision for positives was calculated as the number of TPs divided by the sum of TPs and FPs, whereas precision for negatives was calculated as the number of TNs divided by the sum of TNs and FNs.[34] The F1 score is a measure of predictive performance that incorporates both sensitivity and precision.[35] The F1 score was calculated as in Equation 1.

$$\text{F1score} = \frac{2TP}{(2TP + FP + FN)} \quad (1)$$

Specificity quantifies the accuracy of correctly identifying TNs, and it defines how well dMLA correctly identifies samples that do not contain the target molecules.[34] Specificity was calculated as the number of TNs divided by the sum of TNs and FPs.[34] Because of the imbalanced nature of the data set, which predominantly consists of expected negatives relative to expected positives, the balanced accuracy was also computed.[36] This metric, calculated as the average of sensitivity and specificity, adjusts the TP and TN rates to account for the prevalence of positive and negative cases in the data set. This approach ensures a more reliable measure of performance by giving equal importance to both classes, regardless of their frequency, as shown in Equation 2.

$$\text{Balanced Accuracy} = \frac{\left(\frac{TP}{EP} + \frac{TN}{EN}\right)}{2} \quad (2)$$

Data Availability

All data or links to data as well as all scripts are available at an online repository (*https://github.com/EawagPHH/dMLA*, last accessed March 11, 2025). DNA-sequencing data resulting from the assay are available at Sequence Read Archive (*https://www.ncbi.nlm.nih.gov/sra*, last accessed March 13, 2025) under accession PRJNA1155916. Whole-genome sequencing data are available at DBJ/ENA/GenBank under the accession numbers VNWZ00000000 to VNZG00000000[32] (*https://www.ncbi.nlm.nih.gov/bioproject/527972*, last accessed March 13, 2025) and at the submissions SUB14702901 and SUB14920117 under the BioProject PRJNA1155916 (*https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA1155916*, last accessed March 13, 2025).

Results

Multiplex on Synthetic DNA (Positive Controls)

The multiplex assay was conducted on 63 synthetic DNA templates corresponding to the reverse complementary of each probe-pair included in the multiplex. Each template was used in two technical replicates, and the results from the two replicates yielded consistent outcomes (Figure 1). The thresholds used to differentiate a detect from a nondetect, calculated as the 99.9% quantile of the γ distributed FP results, were applied to filter out background noise in the detection data (Supplemental Table S8). The filtering reduced low-count FPs, refining the detection specificity for each probe-pair and enhancing the clarity of TP identifications (Figure 1).

In the two technical replicates, all 63 probe-pairs successfully identified their corresponding DNA templates with UMI counts substantially exceeding observed UMI counts in samples without the target template (Figure 1, A and B). After filtering the data based on the threshold, the DNA template for each of the 63 probe-pairs exceeded an adjusted UMI count of 1000 (Figure 1, C and D, and Supplemental Table S8). Among the two replicates, only one NTC from the second replicate (Negative4) displayed an FP detection after filtering, with an adjusted UMI count for the probe-pair mcr64 of 1 (Figure 1D). All other NTCs and PCR-negative control across both replicates showed a low number of UMI counts, suggesting probe-pair detection, well below the filter threshold (Figure 1, C and D). Fifty-six probe-pairs were specific and sensitive to their targets, with no FP detections after filtering (Figure 1, C and D). Seven probe-pairs experienced FP detection of off-target synthetic DNA: aac3II63, aac6I123, dfrA369, gadph195, ipaH921, mcr28, and mcr64 (Figure 1, C and D). Notably, these FPs were only observed in one of the two replicates (Figure 1, C and D). The adjusted read count for these FPs was generally low, with 0.6 for the probe-pair aac3II63, 2 for aac6I123, 10 for dfrA369, 0.7 for gadph195, 88 for ipaH921, 12 for mcr28, and 1 for mcr64 (Supplemental Table S9). Notably,

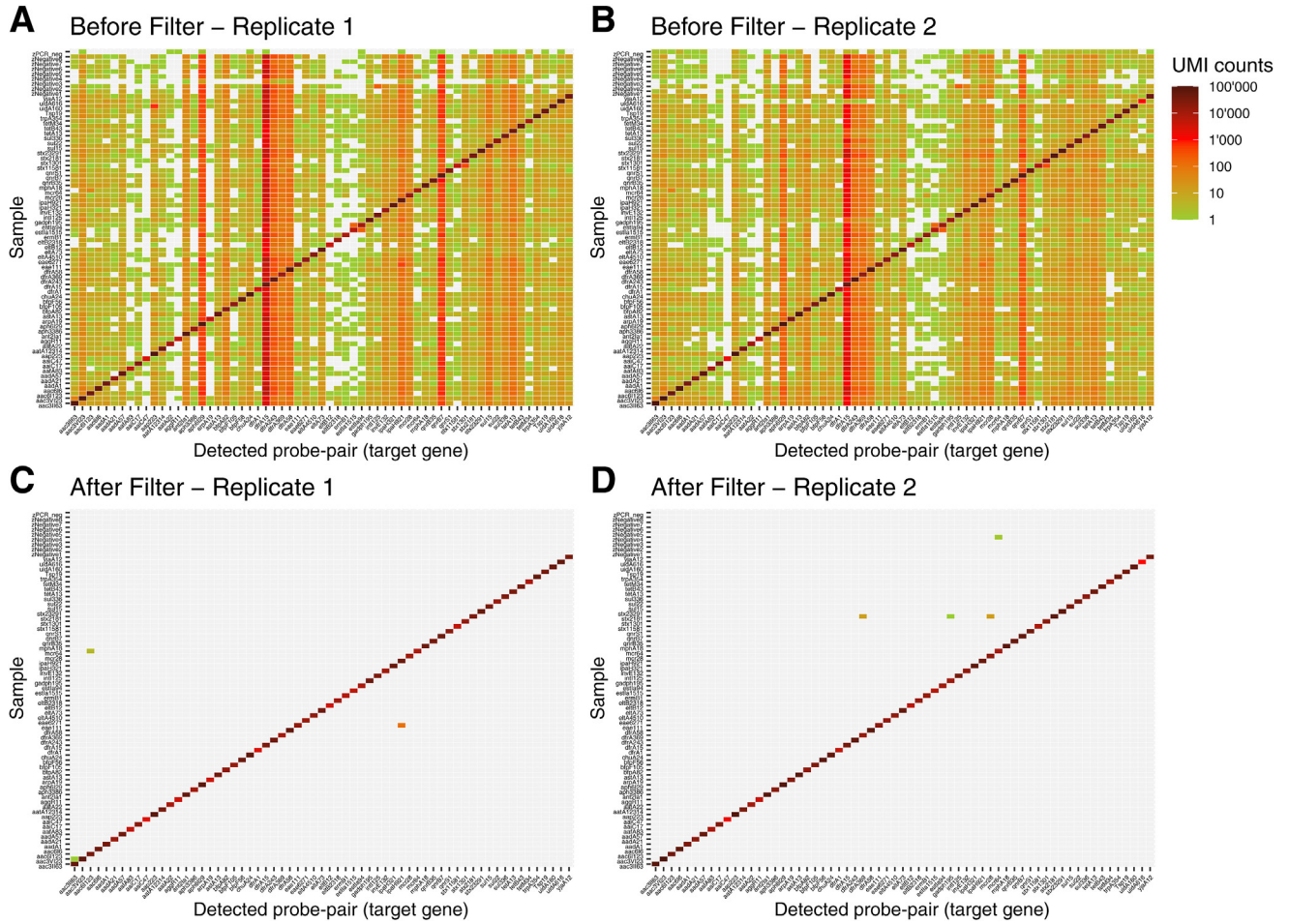

Figure 1 Heat map visualization of unique molecular identifier (UMI) counts from multiplex testing against reverse complementary templates in two experimental replicates. The *y* axis displays the name of the DNA sample tested, either positive template or negative controls. The *x* axis indicates the name of the probe-pairs multiplexed in the assay. Before filter: replicate 1 (**A**) and before filter: replicate 2 (**B**) show the UMI counts for each probe-pair against specific samples before filtering with the threshold. After filter: replicate 1 (**C**) and after filter: replicate 2 (**D**) display the UMI counts after applying a 99.9% quantile threshold of the γ distribution for false-positive mitigation. Each tile corresponds to the read count of a probe-pair, with gray tiles indicating 0 UMI counts, signifying nondetection of the gene target. The color gradient extends from green to red, representing increasing UMI counts, and turns black for UMI counts exceeding 100,000.

samples that were expected to be positive (TP detections) had substantially higher UMI counts: 85,005 for aac3II63, 43,549 for aac6I123, 15,839 for dfrA369, 12,994 for gadph195, 60,893 for ipaH921, 22,605 for mcr28, and 10,835 for mcr64, indicating >800-fold increase relative to the FPs (Supplemental Table S10). FP UMIs were rare and generally showed low adjusted read counts across negative samples, with the filtering approach refined using a γ model based on data from a reaction that included 50 negative controls to accurately capture background noise (Supplemental Figure S2). Finally, technical replicates displayed small variations in probe-pair UMI counts (Supplemental Figure S3 and Supplemental Table S10).

Multiplex on Bacterial Isolates

The multiplexed dMLA assay was conducted on 66 bacterial isolates, including 58 *E. coli*, 2 *S. aureus*, 2 *K. pneumoniae*, 1 *K. oxytoca*, 1 *V. cholera*, 1 *P. lurida*, and 1 *S. enterica* isolates (Figure 2). The assay was applied to each isolate in duplicate across three distinct reactions. The thresholds calculated as the 99.9% quantile of the γ distributed FP results were applied separately to each reaction to filter out background noise in the detection data (Supplemental Table S8).

The assay successfully identified 86% of the expected positives (Figure 2). Correspondingly, 14% of the expected positives were classified as FNs. The assay correctly identified 99% of expected negative and 1% of FP (Figure 2). The expected positives were determined on the basis of the presence of the probe-pairs in the whole-genome sequences of the isolates. In all three reactions, all NTCs and PCR-negative controls were negative (UMI counts below the threshold), and positive controls consistently showed expected positive results in both duplicates (Figure 2 and Supplemental Figure S4).

Figure 2 Each cell represents the assay result for a specific probe-pair (*x* axis) and sample combination (*y* axis). The *y* axis displays the name of the DNA sample tested, which includes *Escherichia coli* isolates (red labels), non–*E. coli* isolates (purple labels), synthetic DNA positive controls (blue labels), and negative controls (green labels). The *x* axis indicates the name of the probe-pairs multiplexed in the assay. The color coding indicates the type of result: false negatives (orange), false positives (yellow), true negatives (gray), and true positives (green). The expected positives were determined on the basis of the presence of the genomic region targeted by the probe-pairs in the whole-genome sequences of the bacterial isolates. Presence of a probe-pair is indicated when unique molecular identifier counts exceed 0 after filtering out the 99.9% quantile threshold of the γ distribution for false-positive mitigation. *K. oxytoca*, *Klebsiella oxytoca*; *K. pneumoniae*, *Klebsiella pneumoniae*; *P. lurida*, *Pseudomonas lurida*; *S. aureus*, *Staphylococcus aureus*; *S. enterica*, *Salmonella enterica*; *V. cholera*, *Vibrio cholera*.

The assay showed high specificity across non−*E. coli* isolates: the *uidA* marker was absent in all tested non−*E. coli* strains (Figure 2 and Supplemental Figure S4). The negative control marker *gapdh*, specific to *Klebsiella* species, was detected only in *K. pneumoniae* and *K. oxytoca*. In one of the two *K. pneumoniae* strains, six probe-pairs (aac3II63, aac6II23, aac6I6, aph6I29, intI125, and tetA13) targeting resistance genes were detected. Probe-pairs aph6I29 and sul22 were also identified in *S. enterica*. No probe-pairs were detected in *S. aureus, P. lurida*, or *V. cholera* isolates.

Of all probe-pairs, bfpF56, bfpF105, aafA83, aap223, aggR11, eltB12, ipaH321, and qrnB35 were not detected in the whole-genome sequences of any bacterial isolate but exhibited at least one FP result (Figure 2). Further BLAST analysis, even with relaxed percentage identity thresholds, confirmed the absence of these probe-pairs in the genomic sequences of all tested isolates. Conversely, probe-pairs dfrA15, dfrA243, dfrA369, eae6271, invE132, and qnrB7 were not detected in any bacterial isolate, despite being expected in specific isolates based on whole-genome sequencing results, leading to FNs (Figure 2). These probe-pairs were characterized by high threshold values for filtering because of substantial UMI counts in the negative controls (Supplemental Table S8), suggesting a lack of specificity for the probe-pairs.

Each bacterial isolate was tested in duplicate within the same reaction, showing highly similar results between duplicates (Figure 2 and Supplemental Table S11). Most samples demonstrated identical outcomes for probe detection in both duplicates, with most achieving 100% agreement. Notably, all negative controls and positive controls (uidA160, aaiC47, and aggR11) also achieved 100% agreement, indicating consistent and reliable results across both duplicates (Supplemental Table S11). However, a few samples showed discrepancies between duplicates, where certain probe-pairs were detected in one replicate but not in the other. The bacterial isolate with the largest discrepancy between duplicates nevertheless had agreement of 91% ($n = 57$) of the probe-pairs within the multiplex, with 9% ($n = 6$) of the probe-pairs exhibiting different outcomes (Supplemental Table S11).

Assay Performance Evaluation

The performance of the multiplex assay was evaluated separately for each technical replicate of synthetic DNA and for the bacterial isolates, revealing notable differences across metrics, such as sensitivity, specificity, balanced accuracy, precision, and F1 score (Table 2).

The assay demonstrated high sensitivity across all three experiments. In both replicates tested on synthetic DNA, the sensitivity was 100%, correctly identifying all expected positives as positive (Table 2). However, sensitivity decreased to 81% when evaluating the assay on bacterial isolates, with fewer anticipated gene targets accurately detected compared with synthetic templates.

The assay exhibited nearly perfect specificity, achieving >99.9% in both replicates tested on synthetic DNA. Almost all expected negatives were correctly identified as negative, with <0.1% of the total tests resulting in FPs. When applied to bacterial isolates, specificity remained high at 99% (Table 2).

Balanced accuracy, estimated as the average of sensitivity and specificity, was high for synthetic and bacterial DNA samples, underlining the robust performance of the assay. Specifically, it exceeded 99.9% in both replicates tested on synthetic DNA, and reached 90% when evaluated on bacterial isolates (Table 2).

Precision, an indication of the proportion of positives that are TPs and of negatives that are TNs, was also high for the assay (Table 2). When tested on synthetic DNA, the precision on positive data was 95% for one replicate (indicating 5% of positives are FPs) and 94% for the second replicate (indicating 6% are FPs). For the bacterial isolates, precision was 86%, indicating that testing on bacterial isolates is associated with higher likelihood of FPs. The precision on negative data was much higher, reaching 100% on both replicates of synthetic DNA and 99% on bacterial isolates.

The F1 score, which is an indicator of balanced performance between sensitivity and precision, was 98% and 97% for the two replicates tested on synthetic DNA (Table 2). As with the other indicators, the F1 score was reduced when evaluating on bacterial isolates to 83%.

Cost-Effectiveness of dMLA for Large-Scale Screening

The cost analysis of the dMLA for varying sample sizes in Switzerland, as of 2024, demonstrates low costs per isolate. The setup costs needed to sequence one or more isolates are approximately US$6131, including expenses for probes, primers, enzymes, and sequencing (Table 3). For 100 isolates, the cost per isolate is, therefore, approximately US$61. As much of the setup costs (ie, probes, primers) can be used to characterize several thousand isolates, the costs per isolate decrease as the number of isolates screened increases. For example, for 1000 isolates, the cost is reduced to US$9 per isolate; for 5000 isolates, it is further reduced to US$5 per isolate; and for 10,000 isolates, the cost per isolate is reduced to US$4. This highlights the low costs achievable for applying dMLA to large libraries of *E. coli* isolates.

Discussion

The dMLA described here exhibits a high multiplexing capacity for targeted detection of gene regions important for the characterization of *E. coli*, including phylogroup, antibiotic resistance, and virulence. This assay builds on the promising efficacy previously reported for the digital multiplex ligation assay method, which is designed to simultaneously screen up to hundreds of bacterial isolates.[21] In this study, a total of 72 samples were screened in a single

Table 2 Predictive Performance Metrics of the Digital Multiplex Ligation Assay Across Synthetic DNA and Bacterial Isolates

Variable	Total number of cases	EP cases	EN cases	TPs	FPs	TNs	FNs	Sensitivity, %	Specificity, %	Precision on positive data, %	Precision on negative data, %	F1 score, %	Balanced accuracy, %
Replicate 1: synthetic DNA	4536	63	4473	63	3	4470	0	100.00	99.93	95.45		97.67	99.97
Replicate 2: synthetic DNA	4536	63	4473	63	4	4471	0	100.00	99.91	94.03		96.92	99.98
Bacterial isolates	13,482	890	12,580	766	124	12,399	181	80.9	99.0	86.1	98.6	83.4	89.9

The table displays total number of possible cases, EP cases, EN cases, TPs, FPs, TNs, FNs, and the calculated values for sensitivity, specificity, precision (on positive and on negative data), F1 score, and balanced accuracy. Sensitivity was calculated as TP/(TP + FN), specificity as TN/(TN + FP), precision as TP/(TP + FP) and TN/(TN + FN), F1 score as 2 TP/(2 TP + FP + FN), and balanced accuracy as (TP/EP + TN/EN)/2.

EN, expected negative; EP, expected positive; FN, false negative; FP, false positive; TN, true negative; TP, true positive.

run, including bacterial isolates and appropriate controls. The assay was further applied to an expanded suite of 19 antibiotic resistance genes, 1 antimicrobial resistance marker (*intI1*), 16 VFs, 5 *E. coli*–specific phylogroup markers, and 2 genes for species-level confirmation of *E. coli* or *K. pneumoniae*. The assay demonstrated high accuracy and performance when tested on both synthetic DNA and a library of DNA extracts of bacterial isolates, where the *uidA* marker or other unintended targets were not detected in non–*E. coli* isolates, and *gapdh* was detected only in *Klebsiella* species, as expected. False positives or misclassification can occur if samples do not represent individual bacterial isolates. For example, if isolation of bacterial cultures leads to enrichment of more than one bacterial isolate (eg, a co-culture of *E. coli* and *Klebsiella* species).

The dMLA is a highly promising method for high-throughput screening, allowing simultaneous detection of multiple gene regions and the screening of hundreds of bacterial isolates in a single run.[21] The ability to pool PCR products from multiple samples, each marked with a unique barcode, significantly reduces sequencing costs, making it an efficient and cost-effective method for extensive sample analysis. The sequencing costs to detect the presence of 43 genes could be reduced to approximately US$4 to US$5 per sample when scaling up to screen 5000 to 10,000 isolates, as of 2024. This is a reagents-only cost analysis and does not include personnel time and equipment costs, which might vary widely across countries. Personnel and equipment costs for the described method are expected to be lower than for whole-genome sequencing, given the streamlined workflows and fewer technical requirements. For comparison, the cost of whole-genome sequencing in different countries around the world ranges from US$72 to US$470 per isolate, as published in a recent review.[37] Moreover, further cost reductions could be achieved by increasing the number of barcoded primers (namely, by pooling together a higher number of samples for sequencing, with the primary limitation theoretically driven only by sequencing depth).[21] In low-and middle-income countries, where screening for pathogenic and antimicrobial resistant markers in bacterial isolates is often limited, the cost-efficiency and high-throughput nature of dMLA can significantly improve the monitoring of pathogenic and antimicrobial resistant *E. coli* in environmental samples.[38]

Implementing dMLA to screen large *E. coli* collections could offer a comprehensive view of prevalent pathogenic and resistant strains, supporting more targeted and regionally effective public health interventions. Such an approach could be particularly beneficial in strengthening programs like the Multiple Indicator Cluster Survey, which assesses

Table 3 Cost Analysis of Digital Multiplex Ligation Assay for Different Sample Sizes

Variable	Maximum isolates per stock (estimated)	Initial costs per stock (estimated), US$	100 Isolates, US$	1000 Isolates, US$	5000 Isolates, US$	10,000 Isolates, US$
Left probes	50,000	575	575	575	575	575
Right probes	50,000	3106	3106	3106	3106	3106
Reverse primer	5000	4	4	4	4	8
Forward primer	5000	1140	1140	1140	1140	2280
Ligase enzyme	600	369	369	738	3321	6273
DNA polymerase	2000	664	664	664	1992	3320
Sequencing (5 million read pairs)	100	274	274	2740	13,700	27,400
Total costs	—	6131	6131	8966	23,837	42,961
Costs/isolate	—	—	61	9	5	4

The table shows the initial setup costs and the resulting cost per isolate for 100, 1000, 5000, and 10,000 samples. The table includes the maximum number of samples that can be processed with one stock of each component. The higher cost of the right probe compared with the left probe is due to 5′ phosphorylation required for ligation. Forward primers are more expensive than reverse primers because they include molecular barcodes for multiplexing. The sequencing cost of US$274 per 100 samples reflects the pooled sequencing of purified PCR products from all samples.

—, Not applicable.

drinking water quality using *E. coli* as an indicator of fecal contamination in low-and middle-income countries, but does not differentiate between pathogenic and resistant strains.[22]

Although the assay only achieved 90% balanced accuracy when tested on bacterial isolates, its primary utility is attributed to the high-throughput screening capability. Lower balanced accuracy indicates that there is a disparity in the detection of TPs and TNs. Specifically, it was found that the assay on bacterial isolates has a higher rate of FNs, for specific probe-pairs more than for others, and it also occasionally produced FPs by incorrectly identifying negative samples as positive.

False negatives might be explained by high UMI counts observed in the negative controls, which caused the estimated thresholds (used to filter reads to differentiate between a detect and nondetect) to also be high. Specifically, when negative samples exhibit high UMI counts for a probe-pair, the threshold for detecting a TP for the corresponding probe-pair will be high. If this threshold is equal to or exceeds the UMI counts in TP samples, it can result in misclassifying these TPs as (falsely) negatives. High UMI counts in negative samples could be due to non-specific binding or cross-reactivity, such as probes binding to other probes during ligation—an affinity not detected *in silico*. A higher rate of FNs was observed among probe-pairs targeting different regions of the same gene, suggesting possible competition during the ligation process or cross-binding in negative samples, which may compromise assay performance by reducing the limit of detection and increasing the probability of misrepresentative results.[39] When specific primer-probes are problematic for FNs (eg, the dfrA369 probes), redesigning to target different gene regions may help to reduce specificity issues in nondetects, or otherwise increase assay performance.

False positives were observed as well, which could be attributed to several factors. False positives with low UMI counts can be influenced by the distribution of UMI counts in the negative samples. The cause of the detection of a target gene in only one NTC in the study (with an adjusted UMI count of one for mcr64) is unknown, but it is likely due to contamination during sample preparation. All NTCs in each replicate were prepared using the same nuclease-free water; if the water was contaminated, it likely would have affected many or all of the NTCs. Additionally, presence of the target gene before PCR amplification would have resulted in significantly higher post-filtration UMI counts, resembling UMI counts observed in TPs with synthetic DNA controls. Other potential sources of the FP cannot be excluded, such as cross-contamination of amplified products after PCR with low levels during or after sequencing steps, or primer-dimer formation. However, this appears to be a stochastic process, as no FPs were detected in the NTCs during bacterial isolate testing. The fact that only 1 of 16 NTCs across two replicates on synthetic DNA showed an FP underlines the random and infrequent nature of such occurrences, aligning with the explanation of thresholding-driven variability in UMI counts. A higher number of negative samples in each assay might provide a more accurate distribution to estimate the threshold for detect/non-detect classification. Ensuring a reliable distribution helps in setting thresholds that minimize the inclusion of FPs. Non-specific amplification and probe binding to nontarget sequences can also contribute to FPs.[40] Although primer-dimer formation was assessed *in silico*, such interactions may not always be detectable computationally. In a complex real matrix, chemical reactions during ligation may still allow for non-specific binding or cross-reactivity, such as probes binding to other probes. Similarly, carry-over contamination or other inefficiencies in the assay process may cause the observed low concentration of target reads in all samples independent from the true presence of the target gene. If the number of these reads is low in DNA samples, even if the number is low in the negative controls, the threshold might not be high enough to filter them out, resulting in low-count FPs. Additionally, laboratory practices, such as cross-contamination during sample processing, can also lead to FPs.[40] Similarly, FPs might also be attributed to background noise introduced during the library preparation for sequencing.

The dMLA assay demonstrated high specificity, with a γ distribution model applied to FP UMIs to accurately set detection thresholds. Adjusting these thresholds further risks misclassifying TPs, particularly for probe-pairs with low UMI counts in positive samples. Notably, the reliability of the filtering threshold increases with the inclusion of a higher number of negative samples, which provides a more accurate estimate of the background distribution.

The FPs reported in this study may also result from imperfect reference genomes to determine the ground truth. In this study, FPs may have resulted from inaccurate non-detection of the gene in the whole-genome sequences. For instance, genes located on plasmids may not be detected if the plasmid is lost during the culturing process or during DNA extraction, or if the plasmid DNA is underrepresented in the sequencing data because of its lower abundance compared with chromosomal DNA.[41] Additionally, plasmid sequences can be more difficult to assemble and identify in whole-genome sequencing data, leading to their absence in the assembled genome but their presence being detected by the assay.[41] Incomplete or fragmented genome assemblies where certain regions are missing or poorly covered can also contribute to the nondetection of these genes in the sequencing data.[42]

Despite these challenges, the potential to multiplex probes for comprehensive characterization of bacterial isolates using dMLA is high, offering a robust first step for high-throughput screening. dMLA may also offer an initial screening approach, which can be followed by PCR confirmation for enhanced reliability or for helping to identify isolates for subsequent whole-genome sequencing. Indeed, in scenarios where extensive sample analysis is

necessary, a balanced accuracy of 90% still confers robust reliability across a high number of bacterial isolates, underlining high throughput over more stringent accuracy detection methods.[43,44]

The more conceptual limitations of dMLA include the requirement for highly specific and sensitive probe-pairs, which can be challenging to develop. This is especially difficult considering the need to account for inter-probe binding using *in silico* methods and to avoid dimer formations. Additionally, the targeted nature of dMLA necessitates predefined probes for specific genes, potentially overlooking untargeted genetic elements. This issue is shared with all PCR-amplification or other enrichment-based methods,[45] but may be overcome by nontargeted sequencing-based methods, such as shotgun metagenomics.[46] Furthermore, mutations within the probe region in samples may reduce or inhibit probe-pair ligation, leading to FNs and impacting assay sensitivity. Another limitation is that dMLA requires sequencing of samples, which, at least with the current sequencing methods, takes more time to prepare than alternative methods, including multiplex PCR using a gel or quantitative real-time PCR using fluorescent probes. Although an isolate is not required for assay development, as the assay can be tested on synthetic DNA, it is necessary for applying the assay to actual samples, which may limit initial high-throughput screening applications. Finally, although the assay demonstrated high specificity, certain VFs, such as *aggR* (enteroaggregative *E. coli*) or *stx1* and *stx2* (enterohemorrhagic *E. coli*), can be present in more than one pathotype, potentially complicating differentiation.

This assay can be expanded to detect additional targets, including other genes in *E. coli* and other microbial targets. Future development could focus on applying dMLA to diverse matrices, including clinical and environmental samples. The potential for quantification based on the number of UMIs also offers exciting opportunities for further research. Reliable quantification with dMLA requires assay optimization to address variations due to gene count differences per cell and inefficiencies in probe binding, ligation, and PCR amplification.[21] Currently, its strength lies primarily in determining gene presence or absence.

The dMLA presented here extends the previous work on extended spectrum β-lactamase genes to target a comprehensive and complete suite of genes to characterize *E. coli*, including phylogroup, VFs, and antimicrobial resistance. These probes were designed using *in silico* methods to enhance target coverage and streamline laboratory procedures, achieving a 100% sensitivity rate in synthetic DNA tests. The enhanced multiplexing capacity with clinically important probes offers a robust tool for high-throughput screening of *E. coli*, facilitating large-scale data collection and enhancing disease surveillance and resource allocation strategies.

Acknowledgments

We thank Lea Caduff for support in the assay development; Tanja Stadler, Antonia Eichelberg, Anne Vescovi, Seju Kang, Melissa Pitton, Esther Greenwood, Tim Vaughan, Louis du Plessis, Jolinda de Korne, Etthel Windels, and James Munday for providing insightful inputs for the discussion; Adrian Lison and Patrick Schmidhalter for supporting the analyses; James Platts-Mills and M. Aminul Islam for guidance in selecting the targets; and Patrice Nordmann's laboratory (University of Fribourg, Fribourg, Switzerland), the Microbial Systems Ecology Group (Eawag, Duebendorf, Switzerland), and Frederik Hammes' laboratory (Eawag, Duebendorf, Switzerland) for providing bacterial isolates.

Author Contributions

S.C. conceptualized the study; developed methods; validated, analyzed, visualized, and curated data; performed experiments; and wrote, reviewed, and edited the manuscript; P.R.O. developed methods, performed experiments, and reviewed and edited the manuscript; M.T. conceptualized the study, obtained software, analyzed data, and reviewed and edited the manuscript; and T.R.J. conceptualized the study, developed methods, provided resources, reviewed and edited the manuscript, supervised the study, administered the project, and acquired funding.

Disclosure Statement

M.T. and T.R.J. hold a patent (patent number CA3072650A1) on a similar method for detection and quantification of genetic targets in human, animal, and environmental samples. The workflow presented in this study is distinct from that covered by the patent, although it is conceptually related.

Supplemental Data

Supplemental material for this article can be found at *https://doi.org/10.1016/j.jmoldx.2025.03.003*.

References

1. Kaper JB, Nataro JP, Mobley HLT: Pathogenic Escherichia coli. Nat Rev Microbiol 2004, 2:123−140
2. Kotloff KL, Nataro JP, Blackwelder WC, Nasrin D, Farag TH, Panchalingam S, et al: Burden and aetiology of diarrhoeal disease in infants and young children in developing countries (the Global Enteric Multicenter Study, GEMS): a prospective, case-control study. Lancet 2013, 382:209−222
3. Nataro JP, Kaper JB: Diarrheagenic Escherichia coli. Clin Microbiol Rev 1998, 11:142−201

4. Cosgrove SE: The relationship between antimicrobial resistance and patient outcomes: mortality, length of hospital stay, and health care costs. Clin Infect Dis 2006, 42:S82−S89. Suppl 2
5. World Health Organization: Global action plan on antimicrobial resistance. Geneva, Switzerland: WHO, 2015. Available at: https://www.who.int/publications/i/item/9789241509763. Accessed March 13, 2025
6. Seale AC, Gordon NC, Islam J, Peacock SJ, Scott JAG: AMR surveillance in low and middle-income settings—a roadmap for participation in the Global Antimicrobial Surveillance System (GLASS). Wellcome Open Res 2017, 2:92
7. World Health Organization: Antimicrobial resistance: global report on surveillance. Geneva, Switzerland: WHO, 2014. Available at: https://www.who.int/publications/i/item/9789241564748. Accessed April 17, 2025
8. Diallo OO, Baron SA, Abat C, Colson P, Chaudet H, Rolain JM: Antibiotic resistance surveillance systems: a review. J Glob Antimicrob Resist 2020, 23:430−438
9. World Health Organization: Global antimicrobial resistance and use surveillance system (GLASS) report: 2022. Geneva, Switzerland: WHO, 2022. Available at: https://www.who.int/publications/i/item/9789240062702. Accessed April 17, 2025
10. World Health Organization: WHO integrated global surveillance on ESBL-producing E. coli using a "One Health" approach: implementation and opportunities. Geneva, Switzerland: WHO, 2021. Available at: https://www.who.int/publications/i/item/9789240021402. Accessed April 17, 2025
11. European Centre for Disease Prevention and Control and World Health Organization Regional Office for Europe: Antimicrobial resistance surveillance in Europe 2023—2021 data. Copenhagen, Denmark: WHO, 2023. Available at: https://www.who.int/europe/publications/i/item/9789289058537. Accessed April 17, 2025
12. World Health Organization: Central Asian and European surveillance of antimicrobial resistance: annual report 2019. Geneva, Switzerland: WHO, 2019. Available at: https://www.who.int/europe/publications/i/item/9789289058179. Accessed April 17, 2025
13. Ashley EA, Recht J, Chua A, Dance D, Dhorda M, Thomas NV, Ranganathan N, Turner P, Guerin PJ, White NJ, Day NP: An inventory of supranational antimicrobial resistance surveillance networks involving low- and middle-income countries since 2000. J Antimicrob Chemother 2018, 73:1737−1749
14. Buckley GJ, Palmer GH: Combating Antimicrobial Resistance and Protecting the Miracle of Modern Medicine. Washington, DC, National Academies Press, 2022
15. Schreiber C, Zacharias N, Essert SM, Wasser F, Müller H, Sib E, Precht T, Parcina M, Bierbaum G, Schmithausen RM, Kistemann T, Exner M: Clinically relevant antibiotic-resistant bacteria in aquatic environments − an optimized culture-based approach. Sci Total Environ 2021, 750:142265
16. Laprade N, Cloutier M, Lapen DR, Topp E, Wilkes G, Villemur R, Khan IUH: Detection of virulence, antibiotic resistance and toxin (VAT) genes in Campylobacter species using newly developed multiplex PCR assays. J Microbiol Methods 2016, 124:41−47
17. Keenum I, Liguori K, Calarco J, Davis BC, Milligan E, Harwood VJ, Pruden A: A framework for standardized qPCR—targets and protocols for quantifying antibiotic resistance in surface water, recycled water and wastewater. Crit Rev Environ Sci Technol 2022, 52:4395−4419
18. Manaia CM: Framework for establishing regulatory guidelines to control antibiotic resistance in treated effluents. Crit Rev Environ Sci Technol 2023, 53:754−779
19. Guo J, Li J, Chen H, Bond PL, Yuan Z: Metagenomic analysis reveals wastewater treatment plants as hotspots of antibiotic resistance genes and mobile genetic elements. Water Res 2017, 123:468−478
20. Ferreira C, Otani S, Aarestrup FM, Manaia CM: Quantitative PCR versus metagenomics for monitoring antibiotic resistance genes: balancing high sensitivity and broad coverage. FEMS Microbes 2023, 4:xtad008
21. Tamminen M, Spaak J, Caduff L, Schiff H, Lang R, Schmid S, Montealegre MC, Julian TR: Digital multiplex ligation assay for highly multiplexed screening of [beta]-lactamase-encoding genes in bacterial isolates. Commun Biol 2020, 3:1−6
22. Khan S, Hancioglu A: Multiple indicator cluster surveys: delivering robust data on children and women across the globe. Stud Fam Plann 2019, 50:279−286
23. Miller WR, Arias CA: ESKAPE pathogens: antimicrobial resistance, epidemiology, clinical impact and therapeutics. Nat Rev Microbiol 2024, 2024:1−19
24. Zhang AN, Gaston JM, Dai CL, Zhao S, Poyet M, Groussin M, Yin X, Li LG, van Loosdrecht MCM, Topp E, Gillings MR, Hanage WP, Tiedje JM, Moniz K, Alm EJ, Zhang T: An omics-based framework for assessing the health risk of antimicrobial resistance genes. Nat Commun 2021, 12:4765
25. Jahantigh M, Samadi K, Dizaji RE, Salari S: Antimicrobial resistance and prevalence of tetracycline resistance genes in Escherichia coli isolated from lesions of colibacillosis in broiler chickens in Sistan, Iran. BMC Vet Res 2020, 16:267
26. Wang Z, Fu Y, Schwarz S, Yin W, Walsh TR, Zhou Y, He J, Jiang H, Wang Y, Wang S: Genetic environment of colistin resistance genes mcr-1 and mcr-3 in Escherichia coli from one pig farm in China. Vet Microbiol 2019, 230:56−61
27. Jiang H, Cheng H, Liang Y, Yu S, Yu T, Fang J, Zhu C: Diverse mobile genetic elements and conjugal transferability of sulfonamide resistance genes (sul1, sul2, and sul3) in Escherichia coli isolates from Penaeus vannamei and pork from large markets in Zhejiang, China. Front Microbiol 2019, 10:1787
28. Chen L, Yang J, Yu J, Yao Z, Sun L, Shen Y, Jin Q: VFDB: a reference database for bacterial virulence factors. Nucleic Acids Res 2005, 33:D325−D328
29. Clermont O, Christenson JK, Denamur E, Gordon DM: The Clermont Escherichia coli phylo-typing method revisited: improvement of specificity and detection of new phylo-groups. Environ Microbiol Rep 2013, 5:58−65
30. Baltazar M, Bourgeois-Nicolaos N, Larroudé M, Couet W, Uwajeneza S, Doucet-Populaire F, Ploy MC, Da Re S: Activation of class 1 integron integrase is promoted in the intestinal environment. PLoS Genet 2022, 18:e1010177
31. Smolander N, Julian TR, Tamminen M: Prider: multiplexed primer design using linearly scaling approximation of set coverage. BMC Bioinformatics 2022, 23:174
32. Montealegre MC, Talavera Rodríguez A, Roy S, Hossain MI, Islam MA, Lanza VF, Julian TR: High genomic diversity and heterogenous origins of pathogenic and antibiotic-resistant Escherichia coli in household settings represent a challenge to reducing transmission in low-income settings. mSphere 2020, 5:1−17
33. Rognes T, Flouri T, Nichols B, Quince C, Mahé F: VSEARCH: a versatile open source tool for metagenomics. PeerJ 2016, 4:e2584
34. Powers DMW: Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. arXiv 2020, [Preprint], https://doi.org/10.48550/arXiv.2010.16061
35. Taha AA, Hanbury A: Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool. BMC Med Imaging 2015, 15:1−28
36. Mower JP: PREP-Mt: predictive RNA editor for plant mitochondrial genes. BMC Bioinformatics 2005, 6:96
37. Price V, Ngwira LG, Lewis JM, Baker KS, Peacock SJ, Jauneikaite E, Feasey N: A systematic review of economic evaluations of whole-genome sequencing for the surveillance of bacterial pathogens. Microb Genom 2023, 9:000947
38. Iskandar K, Molinier L, Hallit S, Sartelli M, Hardcastle TC, Haque M, Lugova H, Dhingra S, Sharma P, Islam S, Mohammed I, Naina Mohamed I, Hanna PA, Hajj SEl, Jamaluddin NAH, Salameh P, Roques C: Surveillance of antimicrobial resistance in low- and middle-income countries: a scattered picture. Antimicrob Resist Infect Control 2021, 10:63

39. Tighe PJ, Ryder RR, Todd I, Fairclough LC: ELISA in the multiplex era: potentials and pitfalls. Proteomics Clin Appl 2015, 9:406
40. Borst A, Box ATA, Fluit AC: False-positive results and contamination in nucleic acid amplification assays: suggestions for a prevent and destroy strategy. Eur J Clin Microbiol Infect Dis 2004, 23: 289−299
41. Arredondo-Alonso S, Willems RJ, van Schaik W, Schürch AC: On the (im)possibility of reconstructing plasmids from whole-genome short-read sequencing data. Microb Genom 2017, 3: e000128
42. Peona V, Blom MPK, Xu L, Burri R, Sullivan S, Bunikis I, Liachko I, Haryoko T, Jønsson KA, Zhou Q, Irestedt M, Suh A: Identifying the causes and consequences of assembly gaps using a multiplatform genome assembly of a bird-of-paradise. Mol Ecol Resour 2021, 21: 263
43. Browne DJ, Kelly AM, Brady JL, Doolan DL: A high-throughput screening RT-qPCR assay for quantifying surrogate markers of immunity from PBMCs. Front Immunol 2022, 13:962220
44. Dreier M, Meola M, Berthoud H, Shani N, Wechsler D, Junier P: High-throughput qPCR and 16S rRNA gene amplicon sequencing as complementary methods for the investigation of the cheese microbiota. BMC Microbiol 2022, 22:1−18
45. Hedges DJ, Guettouche T, Yang S, Bademci G, Diaz A, Andersen A, Hulme WF, Linker S, Mehta A, Edwards YJK, Beecham GW, Martin ER, Pericak-Vance MA, Zuchner S, Vance JM, Gilbert JR: Comparison of three targeted enrichment strategies on the SOLiD sequencing platform. PLoS One 2011, 6:e18595
46. Quince C, Walker AW, Simpson JT, Loman NJ, Segata N: Shotgun metagenomics, from sampling to analysis. Nat Biotechnol 2017, 35: 833−844