A THESIS FOR THE DEGREE OF MASTER OF SCIENCE

Novel Next Generation Sequencing Panel Method for the Multiple Detection and Identification of Foodborne Pathogens

식중독균 다중 검출 및 식별을 위한 새로운 차세대 시퀀싱 패널 방법

February, 2023

Dong-Geun Park

Department of Agricultural Biotechnology
College of Agriculture and Life Sciences
Seoul National University

Novel Next Generation Sequencing Panel Method for the Multiple Detection and Identification of Foodborne Pathogens

Advisor: Ju-Hoon Lee

Submitting a master's thesis of
Science in Agricultural Biotechnology

February, 2023

Department of Agricultural Biotechnology
College of Agriculture and Life Sciences
Seoul National University

Dong-Geun Park

Confirming the master's thesis written by
Dong-Geun Park
February, 2023

Chair	Do Yeop Lee	(Seal)
Vice Chair	Ju-Hoon Lee	(Seal)
Examiner	Ki Won Lee	(Seal)

ABSTRACT

Dong-Geun Park

Department of Agricultural Biotechnology

The Graduate School

Seoul National University

Detecting and identifying the bacterial origin of foodborne pathogen outbreaks is challenging. However, the NGS panel method could potentially be used to efficiently screen and identify the outbreak origin of various bacteria in one reaction. In this study, two sets of new NGS panel primer sets targeting 18 and 13 specific virulence factor genes from (a) *Bacillus cereus*, *Yersinia enterocolitica*, *Staphylococcus aureus, Vibrio cholerae, V. parahaemolyticus,* and *V. vulnificus* (b) five types of pathogenic *Escherichia coli* (enteropathogenic *E. coli* [EPEC], enteroinvasive *E. coli* [EIEC], enterotoxigenic *E. coli* [ETEC], enterohemorrhagic *E. coli* [EHEC], and enteroaggregative *E. coli* [EAEC clinical (EAEC)])), *Listeria monocytogenes*, and *Salmonella enterica* serovar Typhimurium, respectively, were developed and optimized. Singleplex PCR with the primer sets revealed a single PCR amplicon with the expected size, and a subsequent crosscheck and multiplex PCR revealed no interference in the primer set mixture or

pathogenic DNA mixture, thereby confirming the specificity and selectivity of the new primer sets. In an evaluation of the new NGS panel method, six collected agricultural water samples were contaminated with the six selected foodborne pathogens, and six collected fermented food samples were contaminated with the seven selected foodborne pathogens. NGS panel analysis revealed that 18 target genes were multi-detected in one reaction at 10^8 to 10^5 CFUs per target pathogen and 13 target genes were multi-detected in one reaction at 10^8 to 10^7. Interestingly, the average total sequence read counts from the virulence factor genes were positively associated with the CFUs per target pathogen. Although the NGS panel analysis indicated the advantage of multiple pathogen detection in one reaction, relatively low sensitivity and false positive results occurred with few CFUs (dilution factor of 10^5 in agricultural water and 10^6-10^5 in fermented foods) of the target pathogens. To validate the multiple detection and identification results, two sets and three sets of qPCR analyses were independently performed using the same contaminated agricultural water samples and fermented food samples, respectively, and the efficiency and specificity of target pathogen detection and identification were like those in the NGS panel analysis. Indeed, comparative statistical analysis and Spearman correlation analysis revealed that the NGS panel sequence read counts and qPCR cycle threshold (Ct) values were negatively associated, supporting the similarity of the results. To

further improve NGS panel analysis for more rapid and accurate detection and identification, the NGS panel primer sets must be further optimized and real-time NGS sequencing technology should be used. Nevertheless, this study provides new insights into the application of NGS panel analysis for the multiple detection of foodborne pathogens.

***Keywords*: Next-generation sequencing, NGS panel, Foodborne pathogen, Multiple detection, Real-time PCR**

Student Number: 2021-22202

CONTENTS

3. Results

List of Figures

List of Tables

1. Introduction

Foodborne pathogens, including *Escherichia coli* O157:H7, *Salmonella* spp., *Listeria monocytogenes*, *Bacillus cereus*, *Yersinia enterocolitica*, *Staphylococcus aureus*, and *Vibrio* spp., are widespread and frequently cause foodborne diseases. In the USA from 2009 to 2020, 9,720 foodborne pathogen associated disease outbreaks occurred, causing 168,656 illness, 10,983 hospitalizations, and 268 deaths (Lee and Yoon 2021). In South Korea from 2010 to 2018, 2,815 foodborne and waterborne pathogen associated disease outbreaks occurred, increasing the risk of cervical disease (Lee, Yun et al. 2021). To prevent or reduce such serious foodborne disease outbreaks, it is necessary to rapidly detect foodborne pathogens; thus, the development of efficient foodborne pathogen detection methods is essential (Lee, Lee et al. 2001; Zhao, Lin et al. 2014).

Foodborne pathogen detection methods can be divided into four types: (1) culture-based detection, (2) immunological detection, (3) biosensor-based detection, and (4) DNA-based detection. Culture-based detection is the traditional foodborne pathogen identification method; thus, it has a long history and is considered the gold standard (Bhunia 2014). Using this method, viable colony forming units (CFUs) of foodborne pathogens are detected in genus-specific selective media cultures, and live CFUs and cell

numbers can be confirmed in contaminated samples inexpensively and simply (Bolton 1998). However, at least 2–3 days are required to obtain the results of culture-based foodborne pathogen detection tests, and these are followed by biochemical tests, molecular tests, and/or mass spectrometry (Zhao, Lin et al. 2014). Therefore, alternative rapid foodborne pathogen detection methods have been developed. Immunological detection involves the use of an antibody–antigen reaction to detect foodborne pathogens; the methods used include enzyme-linked immunosorbent assays (Fusco, Quero et al. 2011), lateral flow immunoassays (Shi, Wu et al. 2015), and immunomagnetic separation assays (Shim, Choi et al. 2008). Monoclonal or polyclonal antibodies are used for different specificities to detect specific antigens, offering rapid, portable, and economic detection (Umesha and Manukumar, 2018). However, the influence of environmental stress on the antibody leads to low accuracy in immunological detection (Hahn, Keng et al. 2008). Biosensor-based detection was developed to overcome the disadvantages of immunological detection. Specifically, optical piezoelectric biosensors have been developed that provide a wide working range, rapid results, portability, and enhanced detection accuracy and limit of detection (Zhao, Lin et al. 2014). However, biosensor-based detection requires expensive instruments for analysis, compatible computer software, and trained experts; thus, it is not a cost-effective method (Tokarskyy and Marshall, 2008). DNA-based detection

via specific gene-based polymerase chain reaction (PCR) is generally used in foodborne pathogen diagnostics in laboratories (Priyanka, Patil et al. 2016). Because of DNA amplification, conventional PCR in which specific gene-targeting primers are used exhibits high sensitivity up to the femtogram level (Palka-Santini, Cleven et al. 2009). However, this method still requires a time-consuming electrophoresis step for the detection and confirmation of specific genes, and only one gene can be detected in each analysis (Joensen, Scheutz et al. 2014). To overcome the limitations of conventional PCR, real-time PCR or multiplex PCR methods were developed and optimized. Real-time PCR using specific gene-targeting primers and a probe does not require the electrophoresis step, and specific genes can be detected via the fluorescence signal from the probe (Yang, Chen et al. 2015). Determining fluorescence intensity also enables the quantification of DNA concentrations (Liu, Cao et al. 2019). Multiplex PCR can detect a few targeted genes at the same time because a mixture of primer sets is used (Chen, Tang et al. 2012). Combining these advantages, multiplex real-time PCR was developed. Many PCR-based detection kits developed in recent years use multiplex real-time PCR, which achieves rapid and multiple detection with high specificity and sensitivity (Park, Won et al. 2020). Next-generation sequencing (NGS) has enabled the generation of large quantities of DNA sequences in an economical and time-efficient manner (Gupta and Gupta 2020). The most frequently used

NGS sequencers are those from Illumina, which provide the prevailing high-throughput technology with the highest fidelity (Yohe and Thyagarajan, 2017). Although NGS produces massive amounts of DNA sequences in one run, the technology is expensive (De Magalhães, Finch et al. 2010). However, NGS services have been popularized and subject to reduced costs given the continuous development of new technologies such as nanopore (Oxford NanoPore Technologies, UK) sequencing (Vega, Lerneret al. 2016). Given the reduced costs, this NGS sequencing service is now available for use in molecular studies of foodborne pathogens to achieve rapid detection and identification and facilitate microbial genomics, metagenomics, and even shotgun metagenomics analyses (Chung, Kim et al. 2021; Liu, Liu et al. 2021). NGS panels are promising analysis methods with which hundreds to thousands of target gene sequences can be screened at once and many samples can be simultaneously analyzed to detect and identify foodborne pathogens rapidly and efficiently (Ferrario, Lugli et al. 2017).

The NGS panel method was initially evaluated and used in clinical cancer diagnoses and genetically modified organism (GMO) determination. In a previous study, a NGS panel with 13 endometrial cancer gene target primers was developed and evaluated, and 20 randomly chosen cases of patients with endometrial cancer were successfully classified, highlighting the rapid and accurate diagnosis ability of NGS panels (López-Reig,

Fernández-Serra et al. 2019). In another study, a NGS panel with four GMO-related target gene sequences was developed and evaluated using real-time PCR as the control; the NGS panel and real-time PCR provided a 92% GMO detection rate, indicating the reliability of screening performed via this method (Arulandhu, van Dijk et al. 2018). Given the advantages of NGS panels, they have also been evaluated and tested for the multiple detection and determination of various foodborne pathogens. Prior to the use of NGS panels, the detection and identification of foodborne pathogens was conducted using 16S rRNA–based metagenome and random genome sequencing–based shotgun metagenomics approaches (Bridier 2019; Mira Miralles, Maestre-Carballa et al. 2019); however, these detection methods produce an overabundance of sequence information (Zakotnik, Knap et al. 2022). To overcome this problem, NGS panels were developed and evaluated using specific primer sets, generally targeting the virulence factors and antibiotic resistance genes of foodborne pathogens. However, only one NGS panel study has involved the detection and identification of multiple foodborne pathogens; in this study, a species-specific multiplex PCR amplicon was sequenced using an Illumina MiSeq sequencer to a sensitivity of 10^1 CFUs/g (Ferrario, Lugli et al. 2017). This study demonstrates that, compared with metagenome and shotgun metagenomics sequencing, the NGS panel approach achieves rapid and accurate species-specific identification via

the one-time compact NGS sequencing of virulence factors and antibiotic resistance genes. Only one primer set per pathogen was used in this study, and the specificity and sensitivity of the primer sets were not fully evaluated; however, the importance of NGS panel primer set quality and the requirement of multiple primer sets per pathogen should be considered. Indeed, the NGS panel method should be optimized with reliable multiple primer sets.

In the present study, we aimed to optimize the two sets of NGS panel method for the detection and identification of thirteen major foodborne pathogens in South Korea (a) set 1: *Bacillus cereus, Yersinia enterocolitica, Staphylococcus aureus, Vibrio cholerae, V. parahaemolyticus*, and *V. vulnificus* (b) set 2: five types of pathogenic *Escherichia coli* (enteropathogenic *E. coli* [EPEC], enteroinvasive *E. coli* [EIEC], enterotoxigenic *E. coli* [ETEC], enterohemorrhagic *E. coli* [EHEC], and enteroaggregative *E. coli* [EAEC clinical (EAEC)]), *Listeria monocytogenes*, and *Salmonella enterica* serovar Typhimurium. In addition, 1–5 species-specific primer set(s) per target pathogen were designed and evaluated. With these new primer sets, the NGS panel method was tested and evaluated using the six agricultural water and six fermented food samples contaminated with six or seven selected foodborne pathogens, respectively. To verify the sensitivity and accuracy of the NGS panel, associated multiplex real-time PCR was performed as a control and compared with the NGS panel results.

This study provides a novel optimized NGS panel method that achieves the rapid and accurate detection and identification of selected foodborne pathogens in contaminated samples with efficiency, sensitivity, and accuracy. Therefore, this technology could be useful for securing food safety through the prevention of foodborne disease outbreaks via the rapid and accurate detection and identification of foodborne pathogens.

2. Materials and Methods

2.1. Bacterial strains, media, and growth conditions

The bacterial strains, selective and culture media used in this study are listed in Table 1. All bacterial strains were aerobically incubated at 37°C for 18 h. All culture media were purchased from Oxoid (UK), and the agar medium was prepared with 1.8% BACTO Agar (BD, USA).

Table 1. Bacterial strains, culture medium, samples, and sampling locations

Bacterium	Strain	Selective media[a]	Culture media[b]	Reference[c]	Sample	Sampling location
NGS panel set 1						
Bacillus cereus	SG_003	BBC	LB	This study	Seaweed fulvescens	Garak Agricultural and Fisheries Wholesale Market, Seoul
Yersinia enterocolitica	SG_002	CIN	LB	This study	Pollack roe	Garak Agricultural and Fisheries Wholesale Market, Seoul
Staphylococcus aureus	ATCC 23235	-	LB	ATCC	-	-
	Newman	-	LB	ATCC	-	-
	CCARM 3089	-	LB	CCARM	-	-
	SG_001	MSA	LB	This study	Crab	Garak Agricultural and Fisheries Wholesale Market, Seoul
Vibrio cholerae (non-O1-type)	SG_017	TCBS	LB	This study	Octopus	Noryangjin Seafood Wholesale Market, Seoul
vulnificus	SG_012	TCBS	LB	This study	Mussel	Noryangjin Seafood Wholesale Market, Seoul
parahaemolyticus	SG_014	TCBS	LB	This study	Sea urchin	Noryangjin Seafood Wholesale Market, Seoul
NGS panel set 2						
Escherichia coli						
EAEC	NCCP 14039	EMBA	LB	NCCP	-	-
EHEC	SG_006	EMBA	LB	This study	Chicken breast	Shin-won market, Seoul
EIEC	SG_007	EMBA	LB	This study	Pig intestine	Shin-won market, Seoul
EPEC	SG_010	EMBA	LB	This study	Pig intestine	Shin-won market, Seoul
ETEC	SG_009	EMBA	LB	This study	Chicken gizzard	Shin-won market, Seoul
Listeria monocytogenes	SG_004	OA	LB	This study	Cow intestine	Shin-won market, Seoul
Salmonella enterica serovar Typhimurium	SG_011	XA	LB	This study	Cow intestine	Shin-won market, Seoul

[a], MSA: mannitol salt medium; CIN: cefsulodin–irgasan–novobiocin medium; BBC: Brillance *Bacillus cereus* medium; TCBS: thiosulfate–citrate–bile salts–sucrose medium; OA: Oxford agar medium; XA: xylose–lysine–deoxycholate agar medium; EMBA: eosin–methylene–blue agar medium.

[b], LB: Luria–Bertani medium.

[c], ATCC: American Type Culture Collection; CCARM: Culture Collection of Antimicrobial Resistance Microbes; NCCP: National Culture Collection for Pathogens

2.2. Isolation of foodborne pathogens

For the isolation of foodborne pathogens, five seafood samples and four meat samples were collected from Garak Fisheries Wholesale Market (Seoul, Korea), Noryangjin Seafood Wholesale Market, and Garak market (Table 1). After sample collection, 25 g of the collected samples were transferred to a 3M sterilized bag (USA) and suspended with 225 ml of sterilized phosphate-buffered saline buffer. Suspended samples were homogenized using a BagMixer 400 (Interscience, France) with a speed of 4 m/s for 30 s. After homogenization, the samples were serially diluted to 10^{-5}-10^{-6}, plated on selective agar plates specific for each pathogen (Table 1), and incubated as described previously. After incubation, a single colony was picked and streaked on fresh culture medium in an agar plate (Table 1). The selected bacterium was identified using 16S rRNA gene sequencing technology. In particular, pathogenic *E. coli* was further identified using pathogen type-specific gene PCR. The identified bacterium was stored at −80°C in 20% (w/v) glycerol solution.

2.3. DNA extraction

Bacterial genomic DNA was extracted and purified using a Genelix™ Bacterial Extraction Kit (Sanigen, South Korea) according to the manufacturer's instructions. In preparation for NGS panel analysis, total

bacterial DNA was extracted from prepared agricultural water samples and fermented food samples contaminated with the six and seven selected foodborne pathogens, respectively, or agricultural water and fermented food samples free of these pathogens using a QIAamp DNA Stool Mini Kit (Qiagen, USA) according to manufacturer's standard protocol.

2.4. 16S rRNA gene sequencing

All PCRs were performed using a C1000 Touch Thermal Cycler (Bio-Rad, USA). In addition, 16S rRNA gene sequencing was performed for bacterial identification under the following conditions. The PCR mixture (final volume: 25 μl) contained 1 μl of template DNA (40 ng/μl), 0.5 μl of forward and reverse primers [20 μM; 27F, 5′-AGAGTTTGATCCTGGC TCAG-3′; 1492R, 5′- GGTTACCTTGTTACGACTT-3′ (Chen, Lee et al . 2015)], 12.5 μl of BioFACT™ 2X Taq PCR Master Mix (BioFact, South Korea), and 9 μl of molecular water. The PCR conditions were as follows: 1 cycle of 95℃ for 3 min; 35 cycles of 95℃ for 30 s, 60℃ for 30 s, and 72℃ for 30 s; and 1 cycle of 72°C for 5 min. Following PCR, 16S rRNA amplicons were purified using a NICSROprep™ PCR Clean-up S & V Kit (Bionics, South Korea) and sequenced using a 3730xl DNA Analyzer (Thermo Fisher, USA) at Bionics in South Korea according to manufacturer's standard protocols.

2.5. Pathogenic identification of *E. coli* using PCR

Pathogenic identification of *E. coli* was performed using PCR to identify pathogenic types of isolated *E. coli*. The PCR mixture (final volume of 25 μl) contains as same to 16S rRNA gene PCR mixture except for the primer set. Previously developed primer sets, including EAEC (MP2-aggR-F/R), EHEC (MP4-stx1A-F/R), EIEC (MP2-invE-F/R), EPEC (MP3-bfpB-F/R), and ETEC (MP2-LT-F/R), were used to identify pathogenic types of *E. coli* (Müller, Greune et al. 2007). The PCR condition was as follows: 1 cycle of 95℃ for 3 min; 35 cycles of 95℃ for 30 s, 63℃ for 30 s, and 72℃ for 30 s; and 1 cycle of 72°C for 5 min. In verifying the PCR results, agarose gel electrophoresis was performed with 2.5% agarose gel and ethidium bromide (0.2 μg/ml) and each PCR amplicon size was confirmed in the gel using the 100-bp DNA ladder (Bioneer, South Korea) after gel running at 135 V for 20 min.

2.6. Genome sequencing and analysis

For sequencing library preparation with the bacterial genomic DNA, a TruSeq Nano DNA LT Kit (Illumina, USA) was used to add sequencing barcodes to NGS sequencing templates. The sequencing library was then sequenced using an Illumina MiSeq system according to the Illumina MiSeq 2 × 150 bp paired-end run protocol. The qualified sequence reads were

assembled using the Unicycler program (Wick, Judd et al. 2017), and the assembled contigs of each foodborne pathogen were annotated using the NCBI Prokaryotic Genome Annotation Pipeline (Tatusova, DiCuccio et al. 2016). Genomic sequences of foodborne pathogenic bacteria were deposited in GenBank with BioProject accession numbers PRJNA870224 (*E. coli* NCCP 14039), PRJNA882507 (*S. aureus* SG_001), and PRJNA857825 (other pathogenic bacteria).

2.7. NGS panel primer design and optimization

The publicly available complete genome sequences of target pathogens were collected from the GenBank database in the NCBI (www.ncbi.nlm.nih.gov/genbank) and VFDB (B. Liu, Zheng, Zhou, Chen, & Yang, 2022). Comparative pan-genome analysis with the complete genome sequences of other pathogens was performed using the panX program (Ding, Baumdicker et al. 2018) and ANVIO program (Eren, Esen et al. 2015) to identify target pathogen-specific genes. Among the detected pathogen-specific genes, virulence factors and antibiotic resistance genes were primarily considered for selection. New primer sets for the NGS panel were then designed using the sequences of the selected genes and the Primer3 program (Untergasser, Cutcutache et al. 2012) with the following parameters: size: 100–300 bp; GC content: 40%–60%; T_m value: 53°C–60°C; self-

compatibility: ≥4 (Lorenz 2012). After primer set design, the primer stability, e.g., self-binding and dimer formation, and specificity of the primer set to the target pathogen genome sequence was confirmed using Primer3. For NGS panel sequencing analysis, 1–5 gene(s) per target pathogen were selected for primer design. Therefore, a single pathogen had 1–5 specific primer set(s), and each primer set was optimized as previously explained.

2.8. Singleplex PCR and crosscheck PCR

To validate the primer specificity to the target pathogen genome sequence, singleplex and crosscheck PCRs were performed. For singleplex PCR, the PCR mixture (final volume: 25 μl) contained 1 μl of template DNA (4 ng/μl), 0.5 μl of forward and reverse primers (20 μM), and 12.5 μl of KAPA HiFi HotStart ReadyMix (Roche, Germany), and the final volume was adjusted with molecular water. Crosscheck PCR was used to evaluate the selected primer set, and two approaches were taken: (1) a single primer set with the genomic DNA of target strains and (2) primer set(s) (1–5 primer set(s) per reaction) with the genomic DNA of a single target strain. The PCR mixture of the first crosscheck PCR test was prepared with the same composition as that used in singleplex PCR, except for the genomic DNA templates. The test genomic DNA template mixture for the first crosscheck PCR was prepared with the genomic DNA of a target pathogen and other

nontarget pathogens, and the negative control genomic DNA mixture was prepared with the genomic DNA of only the nontarget pathogens. These template DNA mixtures contained 4 ng/μl of DNA per pathogen. The PCR mixture of the second crosscheck PCR test had the same composition as that used in the singleplex PCR, except for the multiple primer sets, which themselves contained 1–5 primer set(s) (20 μM of each) per reaction with a single target strain. The PCR conditions for both the singleplex and crosscheck PCRs were as follows: 1 cycle of 95℃ for 3 min; 35 cycles of 95℃ for 30 s, 60℃ for 30 s, and 72℃ for 30 s; and 1 cycle of 72°C for 5 min. To verify the PCR results, agarose gel electrophoresis was performed with 2.5% agarose gel containing ethidium bromide (0.2 μg/ml), and the size of each PCR amplicon was confirmed in the gel using 100 bp DNA ladder (Bioneer, South Korea) after the gel was run at 135 V for 20 min.

2.9. Multiplex PCR

In addition to singleplex and crosscheck PCRs, multiplex PCR was performed to confirm the specificity of the primer sets in the multiple detection of target pathogens. The multiplex PCR mixture (final volume: 25 μl) contained 1 μl of template DNA (4 ng/μl per pathogen; 9 pathogens in total), 0.5 μl of forward and reverse primer set(s) (1–5 primer set(s); 20 μM of each), and 12.5 μl of KAPA HiFi HotStart ReadyMix (Roche, Germany),

and the final volume was adjusted with molecular water. The test and negative control genomic DNA mixtures were prepared with the same composition as that used in the first crosscheck PCR. The PCR results were verified following the same procedure used in singleplex and crosscheck PCRs.

2.10. Collection of agricultural water samples and the preparation of contaminated water samples with selected pathogens

Six agricultural water samples were collected from a vegetable farm in Hadong-gun, Gyeongsangnam-do, South Korea. Four samples, namely B4GNG1-1 (chive), B4GNG4-2 (chive), B1GNG8-1 (cabbage), and B1GNG8-2 (cabbage), were collected from ground water, whereas two samples, B1GNS10-1 (cabbage) and B1GNS10-2 (chive), were collected from stream water. One liter of each water sample was collected and transferred into a 2-liter sterilized water pack (Worldmedi, South Korea).

For the preparation of the contaminated water samples, nine pathogenic strains were selected as follows: *V. vulnificus* SG_012; *V. parahaemolyticus* SG_014; non-O1-type *V. cholerae* SG_017; *Y. enterocolitica* SG_002; *S. aureus* strains SG_001, ATCC 23235, Newman, and CCARM 3089; and *B. cereus* SG_003 (Table 1). Nine selected foodborne pathogenic strains were separately inoculated into fresh Luria–

Bertani (LB) media and incubated up to 1.0 optical density at a wavelength of 600 nm. Subsequently, the CFUs of each culture was adjusted to 1.0×10^8 CFUs/ml using sterilized LB broth medium. To prepare a single *S. aureus* culture containing four different strains, 25% of each *S. aureus* strain culture was mixed. Each CFU-adjusted culture of a selected pathogen was mixed, and the culture mixture containing six selected pathogenic species was centrifuged at 13,000 rpm for 10 min to harvest the bacterial cell mixture (1.0×10^8 CFUs per pathogen). This mixed cell pellet (6×10^8 CFUs) was resuspended using 250 ml of each collected agricultural water sample. The resuspended bacterial mixture was then serially 10-fold diluted to 6×10^5 CFUs per sample (1×10^5 CFUs per target pathogen in the sample). These serially diluted agricultural water samples (10^8, 10^7, 10^6, and 10^5 CFUs per target pathogen) were used for total bacterial DNA extractions before further NGS panel analysis. This experiment was performed in triplicate.

2.11. Collection of fermented food samples and the preparation of contaminated fermented food samples with selected pathogens

Six fermented food samples (200 g, each sample), including three kimchi samples prepared with different types of vegetable (cabbage, radish, and leaf mustard), and three yogurt samples prepared in different forms (Greek, yogurt, and liquid yogurt), were collected from a market in Seoul,

South Korea. After sample collection, 25 g of the six collected fermented food sample was separately transferred into a sterilized 50-ml conical tube (SPL, USA). In addition, for the preparation of the contaminated fermented food samples, seven pathogens were selected, including EAEC NCCP 14039, EHEC SG_006, EIEC SG_007, EPEC SG_010, ETEC SG_009, *L. monocytogenes* SG_004, and *Salmonella enterica* serovar Typhimurium SG_011 (Table 1). The seven selected pathogens were separately inoculated into sterilized LB culture media and incubated up to 1.0 optical density at a wavelength of 600 nm. Then, the CFU of each culture was adjusted to 1.0×10^8 CFU/ml with sterilized LB broth medium. Each CFU-adjusted culture of a selected pathogen (1.0×10^8 CFU/ml per pathogen) was mixed, and the mixture containing the six selected pathogenic species was centrifuged at 13,000 rpm for 10 min to harvest the bacterial cell mixture (1.0×10^8 CFU per pathogen). This mixed cell pellet (7×10^8 CFU) was resuspended with 1 ml of PBS (Difco). Then, the resuspended bacterial mixture was serially diluted 10-fold up to 7×10^5 CFU per sample (1.0×10^5 CFU per target pathogen in the sample). These serially diluted bacterial mixture (10^8, 10^7, 10^6, and 10^5 CFUs per target pathogen) was transferred to a 50 ml conical tube (SPL, Korea) containing 25 g of collected sample. After contamination, each of the six contaminated fermented food samples was transferred into a sterilized 3M bag containing 224 ml of PBS (Difco) and homogenized as

previously described. The homogenized samples were used for total bacterial DNA extraction before NGS panel analysis. This experiment was performed in triplicate.

2.12. NGS panel analysis

To prepare the NGS panel sequencing template DNA via PCR, two sets of template DNA were prepared: (a) total DNA for test samples from one of six agricultural water samples or six fermented food samples containing target pathogens and (b) total DNA for negative controls from one of six agricultural water samples or six fermented food samples without the contamination of target pathogens. The PCR mixture (final volume: 25 μl) contained 1 μl of template DNA (the total DNA template for test samples or total DNA template for negative controls), 0.1 μl of forward and reverse primers per primer set (18 primer sets; 100 μM of each), and 12.5 μl of KAPA HiFi HotStart ReadyMix (Roche), and the final volume was adjusted with molecular water. The PCR conditions used were the same as those used in singleplex PCR. Following PCR, target PCR amplicons were gel-extracted and purified using a NICSROprep™ DNA Gel Extraction S & V Kit (Bionics) according to the manufacturer's standard protocol. To prepare the sequencing library, a TruSeq Nano DNA LT Kit (Illumina, USA) was used to add sequencing barcodes to NGS sequencing templates. Subsequently, the

sequencing library was sequenced using an Illumina MiniSeq system according to the Illumina MiniSeq 2 × 150 bp paired-end run protocol. After NGS sequencing, the following steps were taken: (1) a filtering step in which the raw reads were filtered using Trimmomatic (Bolger, Lohse et al. 2014) to obtain a Phred quality score of >20; (2) a merging step in which the filtered reads were merged using Pandaseq (Masella, Bartram et al. 2012) with its default parameters; and (3) a mapping step in which the merged reads were mapped to the six selected pathogen-specific gene sequences using BLASTN with a >95% nucleotide identity (Camacho, Coulouris et al. 2009). Finally, the number of mapped reads was counted. To determine the false positive detection of the NGS panel, it was necessary to determine the detection criteria. To clarify these criteria, NGS panel analyses with six collected agricultural water samples containing or not containing the selected pathogens were conducted and compared.

2.13. Quantitative real-time PCR (qPCR)

To evaluate the NGS panel analysis, quantitative real-time PCR (qPCR) was performed, and the results were compared with the NGS panel analysis results. The qPCR was performed using a CFX96 deep-well plate reader (Bio-Rad). The five sets of NGS panel sequencing template DNA previously described were used as the template DNA for qPCR. A Genelix^{TM}

Multiplex Real-Time PCR Kit (#G102, Sanigen) was used to detect *V. vulnificus, V. parahaemolyticus*, and *V. cholera*, a Genelix™ Multiplex Real-Time PCR Kit (#G104, Sanigen) was used to detect *B. cereus, Y. enterocolitica*, and *S. aureus*, a Genelix™ Multiplex Real-Time PCR kit (#G103, Sanigen, South Korea) was used to detect *L. monocytogenes* and *Salmonella* spp, a Genelix™ Multiplex Real-Time PCR kit (#G105, Sanigen) was used to detect EHEC and ETEC, and a Genelix™ Multiplex Real-Time PCR kit (#G106, Sanigen) was used to detect EAEC, EIEC, and EPEC. The qPCR was performed according to the manufacturer's standard protocols, and the Ct was determined automatically using CFX Manager Software version 3.1 (Bio-Rad). All tests were performed in triplicate.

2.14. Statistical analysis

GraphPad version 7.0 (Prism, USA; http://www.graphpad.com) and R version 4.1.2 (R Core Team 2021) were used to perform all correlations and visualizations.

3. Results

3.1 NGS panel set 1: multiple detection and identification of foodborne pathogens in agricultural water

3.1.1. Isolation and identification of foodborne pathogens

In total, 54 pathogenic bacteria were isolated from 6 seafood samples (crab, pollack roe, seaweed fulvescens, octopus, mussel, and sea urchin). These pathogens were identified as *V. vulnificus* (1 strain), *V. parahaemoly ticus* (1 strain), non-O1-type *V. cholerae* (1 strain), *Y. enterocolitica* (1 strai n), *B. cereus* (1 strain), *S. aureus* (1 strain), *Pseudomonas aeruginosa* (16 st rains), *E. coli* (8 strains), *Klebsiella pneumonia* (2 strains), *Enterococcus hirae* (9 strains), *Enterococcus faecalis* (3 strains), *Listeria innocua* (4 strains), and *Serratia liquefaciens* (6 strains) at the molecular level using 16S rRNA gene sequencing. Among these pathogens, six strains of *B. cereus*, *Y. enterocolitica*, *S. aureus, V. cholera, V. parahaemolyticus,* and *V. vulnificus* were selected as target pathogens, and three *S. aureus* type strains were also selected (Table 1) as these bacterial species have been associated previously with agricultural water contamination related to potential foodborne disease outbreaks (Pianetti, Sabatini et al. 2004; Silva, Caniça et al. 2020; Elshikh, Alarjani et al. 2022; Roulová, Moťková et al. 2022).

3.1.2 General genome features of selected foodborne pathogens and the design of primer sets

The genome sequence information of selected target pathogens is required to design specific primer sets and confirm their binding sites in the genomes. Therefore, NGS genome sequencing was performed, and draft genome sequences were obtained for *B. cereus*, *Y. enterocolitica*, *V. cholera, V. parahaemolyticus,* and *V. vulnificus* as well as two *S. aureus* strains (SG_001 and CCARM 3089). In addition, the previously reported genome sequences of two *S. aureus* strains (ATCC 23235 and Newman) were obtained from the NCBI GenBank database. The general genome features of these foodborne pathogens are summarized in Table 2. Based on the genome sequences, primer sets targeting specific toxin genes and virulence factors were designed to meet the criteria of primer design given in the Materials and Methods. The sequence information of the designed primer sets is shown in Table 3, and the primer target genes and primer binding locations are listed in Table 4.

Table 2. General genome features of selected foodborne pathogens isolated from seafood samples

Bacterium	Strain	Genome size (bp)	Assembly	Contig	GC (%)	CDS	tRNA	rRNA	Reference[d]
Selected foodborne pathogens									
Bacillus cereus	SG_003	5,908,983	Draft	83	34.81	5,920	62	3	This study
Yersinia enterocolitica	SG_002	4,357,829	Draft	123	46.92	3,932	66	3	This study
Staphylococcus aureus	ATCC 23235[a]	2,789,574	Draft	2	32.68	2,705	59	19	ATCC
	Newman[b]	2,878,897	Complete	1	32.89	2,851	59	16	ATCC
	CCARM 3089[c]	2,865,317	Draft	54	32.72	2,822	56	2	CCARM
	SG_001	2,944,975	Draft	111	32.77	2,781	59	4	This study
Vibrio cholerae	SG_017	4,005,842	Draft	91	47.52	3,592	69	4	This study
parahaemolyticus	SG_014	6,040,036	Draft	81	44.01	5,740	135	5	This study
vulnificus	SG_012	5,012,927	Draft	114	46.66	4,401	83	4	This study

a , NCBI GenBank BioProject accession number, PRJNA224116

b , NCBI GenBank BioProject accession number, PRJDA18801

c , NCBI GenBank BioProject accession number, PRJNA870224

d , ATCC, American Type Culture Collection; CCARM, Culture Collection of Antimicrobial Resistance Microbes

Table 3. Selected pathogen species-specific genes, their functions, and the associated designed primer sets in NGS panel set 1

Bacterium	Gene	Function	Primer	Sequence (5' to 3')	Size (bp)	Reference
Bacillus cereus	*entFM1*	Enterotoxin	ent_F	GAACTGCTGGTACAACACCTG	229	This study
			ent_R	TCTGCACTAATGAACTGACCG		
	tpi	Triose phosphate isomerase	tpi_F	GCGCTCTTCTAAAGTCTCAC	175	This study
			tpi_R	CGAAATTAGCCCAGTAGCAC		
Yersinia enterocolitica	*ail*	Attachment invasion locus protein	ail_F	TGGGGCCATCTTTCCGCATTA	235	This study
			ail_R	TACCCTGCACCAAGCATCCAA		
	gspE	Type II secretion system ATPase	gspE_F	AACGGGGCATCTGGTTCTCTC	190	This study
			gspE_R	TGGTGGTGTCAGGAAAGGGAC		
Staphylococcus aureus	*femA*	Methicillin resistance factor	femA_F	GCAGCTTGCTTACTTACTGCT	214	This study
			femA_R	TACCTGTAATCTCGCCATCAT		
	sea1	Exotoxin A	sea_F	ATTCATTGCCCTAACGTGGAC	191	This study
			sea_R	GCTGTAAAAATTGATCGTGACTCTC		
	seb1	Exotoxin B	seb_F	GTATGGTGGTGTAACTGAGC	212	This study
			seb_R	CCGTTTCATAAGGCGAGTTG		
	sec1	Exotoxin C	sec_F	CTGCTATTTTTCATCCAAAGA	180	This study
			sec_R	TTCTTATCAGTTTGCACTTCA		
	sed1	Exotoxin D	sed_F	TGTCACTCCACACGAAGGTA	162	This study
			sed_R	TGCAAATAGCGCCTTGCTTG		
Vibrio cholerae	*ctxA*	Enterotoxin	ctxA_F	GCCAAGAGGACAGAGTGAGTA	253	This study
			ctxA_R	ATGAGGACTGTATGCCCCTA		
	hlyA	Cytolysin and hemolysin	hlyA_F	GTTTGTATGTGCGAGCGGGTG	175	This study
			hlyA_R	GTGAATGTCAGCGCCACCAAC		
	toxS	Transmembrane regulator	toxS_F	TAAGACCAACAGCAACCGCCC	209	This study
			toxS_R	ACTCGACTGGCGTAACCAAAAGG		
Vibrio parahaemolyticus	*plsX*	Phosphate acyltransferase	plsX_F	GCACTGTCTCATTTCCCAGAG	219	This study
			plsX_R	CGCTTCTTGGTCAGAAACCAG		
	tdh	Thermostable direct hemolysin	tdh_F	TCCATCTGTCCCTTTTCCTGCC	187	This study
			tdh_R	CAGCCATTTAGTACCTGACGTTGTG		

	tlh	Thermolabile hemolysin precursor	tlh_F tlh_R	GCGAGCGATCCTTGTTTGGAC GCGGTGAGTTGCTGTTGTTGG	144	This study
	toxR	Transcriptional activator	toxR_F toxR_R	ACCTGTGGCTTCTGCTGTG CCAGTTGTTGATTTGCGGGTG	178	This study
Vibrio vulnificus	*glnA*	Glutamate ammonia ligase	glnA_F glnA_R	AGCACATCTCTATTCCTTCTC TAGCGTTGCTTCTTCAGTAA	170	This study
	vvh	Hemolysin	vvh_F vvh_R	CTCTGCCTAGATGTTTATGG CAATACCATTTCTGTGCTAAG	199	This study

Table 4. *In silico* prediction of primer binding sites in selected foodborne pathogens isolated from seafood samples

Bacterium	Strain	Gene	Contig	Primer binding site	
				Forward (nt)	Reverse (nt)
Selected foodborne pathogens					
Bacillus cereus	SG_003	*tpi*	Contig 4	148,119-148,138	147,964-147,983
		entFM1	Contig 12	62,720-62,740	62,512-62,532
Yersinia enterocolitica	SG_002	*ail*	Contig 62	1,310-1,330	1,524-1,544
		gspE	Contig 2	106,645-106,665	106,476-106,496
Staphylococcus aureus	ATCC 23235	*femA*	Contig 1	1,365,336-1,365,356	1,365,529-1,365,549
		sed1	Contig 2	27,108-27,127	26,966-26,958
	Newman	*femA*	Contig 1	1,417,190-1,417,210	1,417,383-1,417,403
		sea1	Contig 1	2,094,921-2,094,941	2,094,751-2,094,775
	CCARM 3089	*femA*	Contig 10	15,889-15,919	15,706-15,726
		sec1	Contig 19	581-601	740-760
	SG_001	*femA*	Contig 6	15,907-15,927	16,100-16,120
		sea1	Contig 39	410-430	576-600
		seb1	Contig 5	133,856-133,875	133,664-133,683
Vibrio cholerae	SG_017	*ctxA*	Contig 50	1,263-1,283	1,031-1,050
		hlyA	Contig 18	52,696-52,716	52,542-52,562
		toxS	Contig 5	110,538-110,558	110,350-110,372
parahaemolyticus	SG_014	*plsX*	Contig 6	263,592-263,612	263,790-263,810
		tdh	Contig 61	541-561	373-397
		tlh	Contig 1	92,117-92,137	91,994-92,104
		toxR	Contig 13	142,352-142,370	142,509-142,529
vulnificus	SG_012	*glnA*	Contig 18	28,556-28,568	28,417-28,436
		vvh	Contig 11	137,880-137,899	137,682-137,702

3.1.3 Validation of designed primer sets

3.1.3.1. Singleplex PCR

To evaluate the specificity of primer sets to the target pathogens, singleplex PCR was performed with a single target pathogen and an associated single primer set. For the target pathogens, the selected specific genes with their encoded functions, designed specific primer sets, and expected PCR amplicon sizes are listed in Table 3. Following singleplex PCR, agarose gel electrophoresis analysis revealed that all PCR amplicons were of the expected size according to the single PCR bands, confirming the specificity of all the PCR primer sets to the associated target pathogens (Fig. 1). Thus, the designed primer sets qualified for crosscheck PCR evaluation in the next stage.

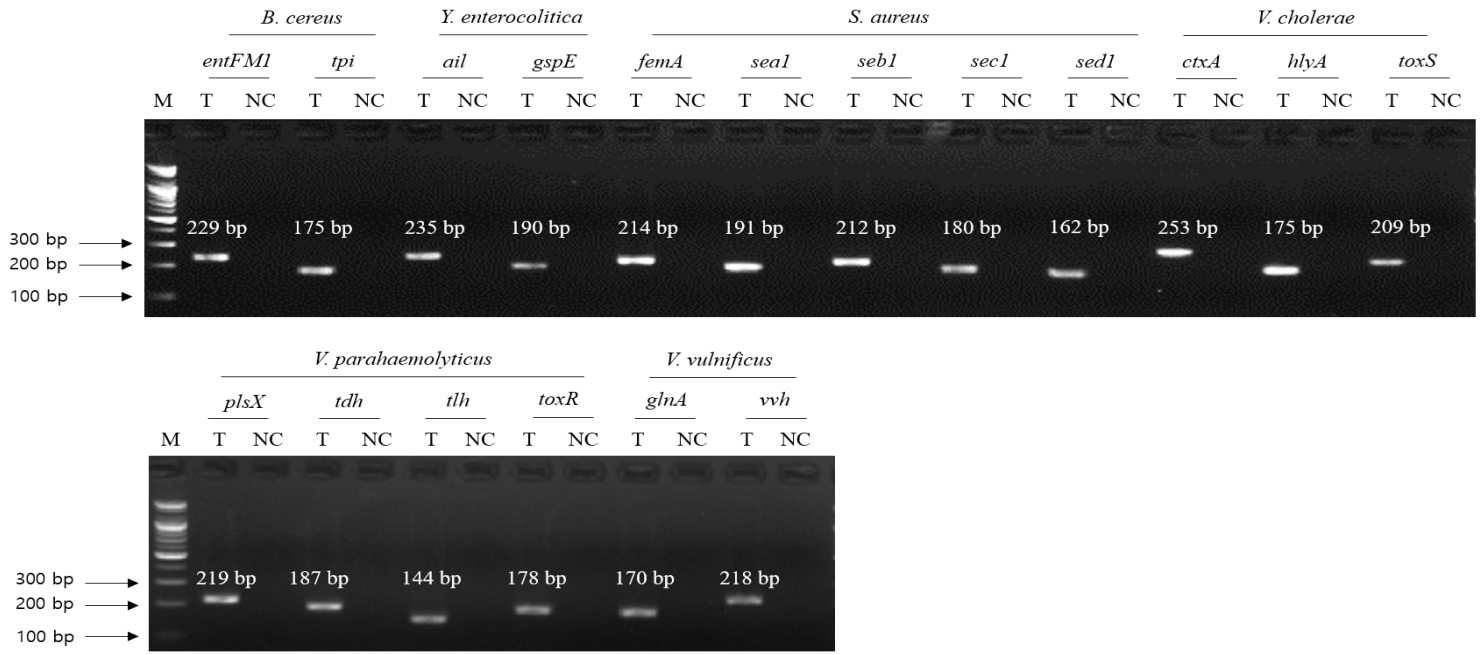

Fig. 1. Gel electrophoresis results of singleplex PCR using NGS panel set 1 primer sets.

Target pathogens and their specific genes are shown above the gel electrophoresis results. PCR mixture of the test (T) lane contained the associated target pathogen genomic DNA and specific gene primer set. PCR mixture of the negative control (NC) lane contained molecular water and the target pathogen-specific gene primer set. M: 100 bp DNA ladder.

3.1.3.2. Crosscheck PCR

To confirm the specificity of the primer sets to the target and nontarget pathogens, two different crosscheck PCRs were conducted: (a) an evaluation of the target pathogen and eight different nontarget pathogens with a single primer set and (b) an evaluation of a single pathogen-targeting primer set mixture (2–5 primer sets) with an associated target pathogen.

For the first crosscheck PCR, two genomic DNA template sets (test DNA template mixture and negative control DNA template mixture without target pathogenic DNA) were prepared to confirm the nonspecific binding of a selected single primer set to nontarget pathogenic DNA. In this crosscheck PCR, the PCR amplicon bands specific to the selected gene were found in the target pathogen but not the nontarget pathogens (Fig. 2). In addition, the sizes of the PCR amplicon bands matched those expected, indicating that the primer sets were highly specific to the target pathogenic DNA, even though the template DNA mixture contained all other nontarget pathogenic DNA. Thus, the PCR primer sets were specific to the associated target gene as well as the target pathogen.

The second crosscheck PCR was conducted to determine whether one PCR reaction can multidetect the target genes in a single pathogen with a single pathogen-targeting primer set mixture that combines primer sets targeting 2–5 selected genes in a single pathogen (Table 3). In this crosscheck

PCR, the PCR amplicons of all target genes in each pathogen were confirmed in the gel electrophoresis (Fig. 3), and their amplicon sizes matched those expected. Therefore, PCR with a mixture of primer sets detected target genes in one reaction without any primer interference.

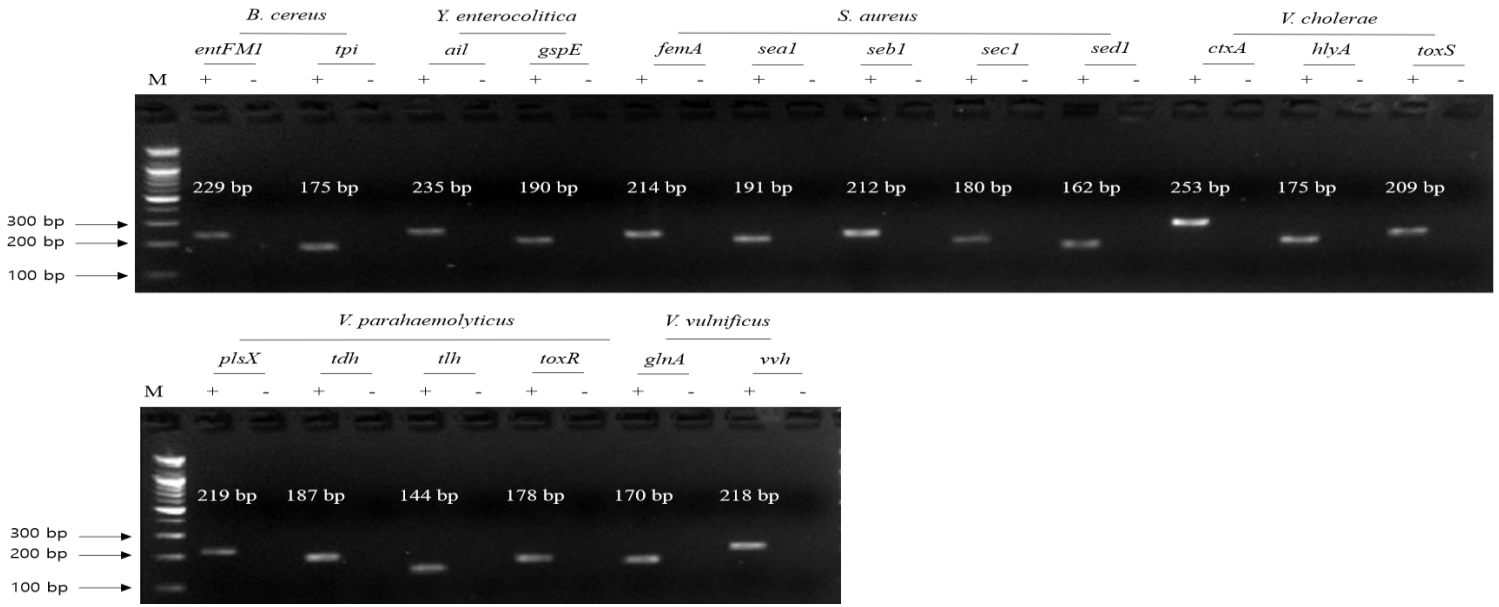

Fig. 2. Gel electrophoresis results of the first crosscheck PCR using NGS panel set 1 primer sets.

Target pathogens and their specific genes are shown above the gel electrophoresis results. PCR mixture of the test (+) lane contained a genomic DNA mixture including the associated target pathogen and target pathogen-specific gene primer set. PCR mixture of the negative test (−) lane contained a genomic DNA mixture lacking the associated target pathogen and target pathogens-specific gene primer set. M: 100 bp DNA ladder.

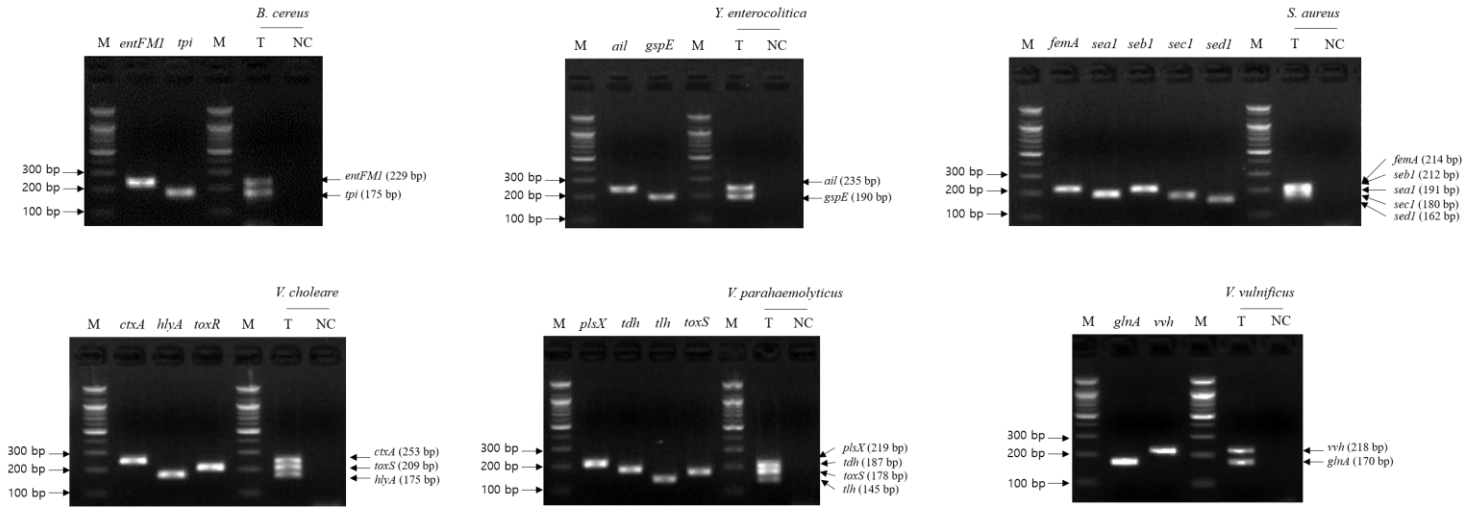

Fig. 3. Gel electrophoresis results of the second crosscheck PCR using NGS panel set 1 primer sets.

Target pathogens and their specific genes are shown above the gel electrophoresis results. Lanes contain each target pathogen-specific gene with singleplex PCR amplicons as positive controls. PCR mixture of the test (T) lane contained the associated target pathogen genomic DNA and 2–5 target pathogen-specific gene primer sets. PCR mixture of the negative control (NC) lane contained molecular water and 2–5 target pathogen-specific gene primer sets. M: 100 bp DNA ladder.

3.1.3.3. Multiplex PCR

To confirm that multiple target genes can be detected in one PCR reaction, multiplex PCR was also performed with the primer set mixture and all target pathogenic DNA. For multiplex PCR, template DNA was prepared with the same sets used in the first crosscheck PCR, and the mixture of primer sets was the same as that used in the second crosscheck PCR. Multiplex PCR results showed that the PCR amplicons of all target genes in each pathogen were detected in gel electrophoresis, and their band sizes were the same as those expected (Fig. 4). Therefore, multiplex PCR confirmed that a mixture of primer sets can be used to detect target genes in one reaction without any template DNA or primer interference.

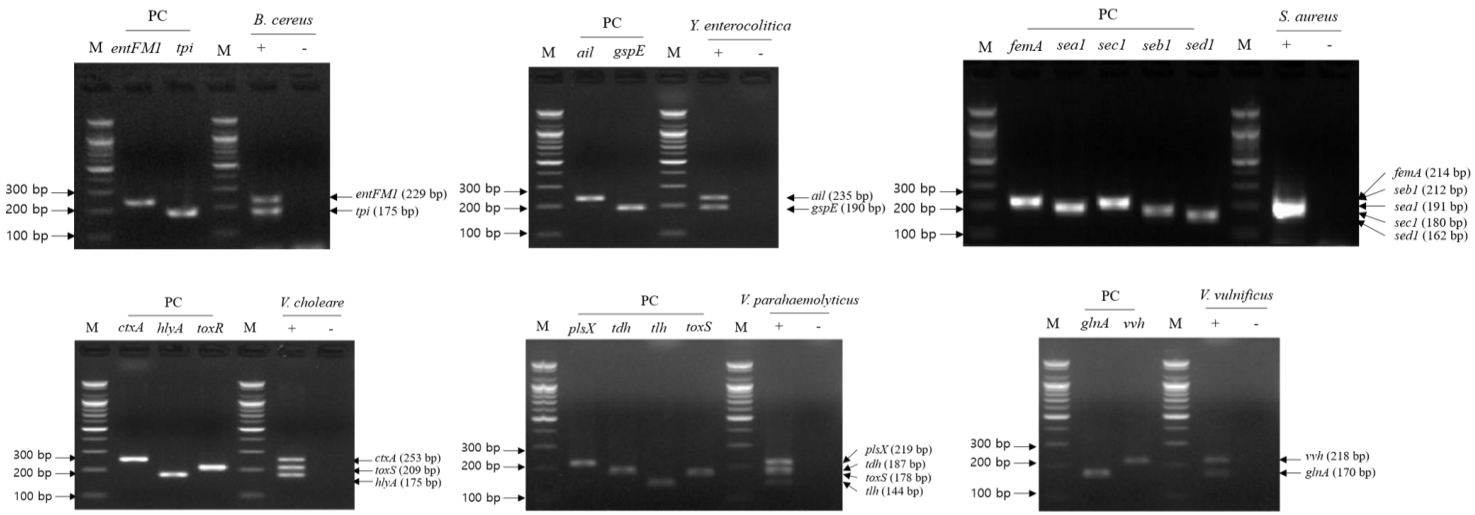

Fig. 4. Gel electrophoresis results of multiplex PCR using NGS panel set 1 primer sets.

Target pathogens and their specific genes are shown above the gel electrophoresis results. Lanes contain each target pathogen-specific gene with singleplex PCR amplicons as positive controls. PCR mixture of the test (+) lane contained a genomic DNA mixture including the associated target pathogen and 2–5 target pathogen-specific gene primer sets. PCR mixture of the negative test (−) lane contained a genomic DNA mixture lacking the associated target pathogen and 2−5 target pathogens-specific gene primer sets. M: 100 bp DNA ladder.

3.1.4 NGS panel analysis

NGS panel analysis was performed with six different agricultural water samples contaminated with a mixture of target pathogens. Following NGS panel sequencing, the qualified sequence reads were collected and mapped to the target pathogen-specific gene sequences. The average number of sequence reads mapped to target pathogen-specific genes was 228,915 (93.263% of total qualified sequence reads), 125,902 (61.501%), 35,360 (23.125%), and 3,218 (1.879%) at dilutions of 10^8, 10^7, 10^6, and 10^5 CFUs per target pathogen, respectively (Table 5 and Fig. 5). Interestingly, the averages number of sequence reads mapped to target pathogen-specific genes and the CFU number per target pathogen were positively associated ($y = 78011x - 95630$; $R^2 = 0.9532$; Fig. 5). The prepared negative control samples without specific pathogen contamination exhibited 1–6 sequence reads mapped to target pathogen-specific genes, suggesting that a small number of pathogens were present in the original agricultural water samples and produced false positive results (Fig. 7A). Thus, ≤6 reads was determined as the false positive rate for further NGS panel analysis. After mapping to 18 different target genes of six target pathogens, all qualified NGS panel sequence reads were collected from the six different agricultural water samples. The collected read counts for each dilution factor (10^8, 10^7, 10^6, and 10^5 CFUs per target pathogen) were compared in terms of the detection and

identification of specific target pathogens (Fig. 6). For the dilution factors 10^6 to 10^8, all 18 target genes were multi-detected, and the dilutions were enough to identify the six target pathogens in one NGS panel analysis without false positives (Figs. 7B-D). This result was confirmed in triplicate tests of all agricultural water samples. As expected, the serial dilution of target pathogens was proportionally associated with the read count, i.e., the highest and lowest numbers of read counts were associated with the dilution factors 10^8 and 10^6, respectively, supporting the results shown in Fig. 5. However, when the dilution factor was 10^5, many false positive results were detected (Fig. 7E). In particular, *tlh* of *V. parahaemolyticus* and *seb1* of *S. aureus* were poorly detected by NGS panel analysis. The number of read counts for each target gene was compared among dilution factors, and the numbers of *tlh* and *seb1* were always lower than those of the other 16 target genes (Fig. 8), supporting the results shown in Fig. 6. Therefore, *tlh* and *seb1* could be removed as target genes to increase the limit of detection and improve the identification of specific target pathogens in NGS panel analysis. Although false positive results were found at 10^5, using the average of triplicate tests in NGS panel analysis removed most of the false positive results at this dilution, thereby enhancing the detection and identification of all target pathogens (Fig. 7E). Nevertheless, the results of NGS panel analysis suggest that the limit of detection and identification of target

pathogens may be at a dilution of 10^5 CFUs.

Table 5. Summary of NGS panel set 1 output in six agricultural water samples with or without target pathogen

Sample	Replicate	CFU	Yield (bp)	Raw read	Filtered read	Merged read	Mapped read to total target pathogens specific genes
B4GNG1-1	1	10^8	122,323,563	813,624	777,011	388,092	225,870
	2		120,135,561	798,876	763,566	382,365	226,278
	3		113,298,422	753,494	718,909	360,664	210,396
	1	10^7	102,606,782	681,612	651,757	324,788	117,955
	2		102,524,845	681,180	650,391	322,964	133,514
	3		88,639,999	589,026	564,287	283,870	87,610
	1	10^6	106,108,444	705,190	669,155	323,990	84,502
	2		103,007,095	684,696	649,023	310,960	75,390
	3		86,276,123	573,070	547,053	269,964	44,552
	1	10^5	72,610,739	483,468	443,534	226,855	2,094
	2		77,685,911	515,642	490,066	244,924	3,782
	3		80,914,390	536,924	515,769	260,710	2,229
	1	N.C.	66,054,304	438,868	424,824	190,947	0
	2		65,853,979	437,844	423,527	211,562	0
	3		59,185,593	393,074	379,788	211,568	0
B1GNG8-1	1	10^8	122,323,563	813,624	687,548	343,648	215,557
	2		113,840,513	756,878	724,938	362,091	214,609
	3		93,555,990	622,006	595,322	299,075	200,636
	1	10^7	99,330,855	660,706	634,146	317,345	140,911
	2		74,456,985	494,550	474,175	238,829	106,379
	3		99,520,131	661,242	634,726	319,419	166,740
	1	10^6	75,483,859	501,438	480,227	241,545	51,230
	2		68,273,889	452,794	434,637	219,744	7,581
	3		73,256,917	486,380	467,071	236,470	26,051
	1	10^5	63,445,162	420,648	399,868	204,908	2,494
	2		61,515,362	408,040	381,517	195,097	9,668
	3		58,458,568	387,580	365,488	188,214	542
	1	N.C.	65,557,295	435,878	420,622	197,030	0
	2		59,703,457	396,256	382,110	193,337	1
	3		62,023,488	412,520	397,587	209,078	1
B1GNS10-1	1	10^8	123,664,901	821,608	784,061	392,846	241,428

	2		116,073,213	772,092	738,043	366,528	222,473
	3		124,383,663	826,410	790,131	397,656	276,753
	1	10^7	84,073,800	558,692	532,378	256,122	98,591
	2		87,042,497	579,368	551,385	265,964	91,250
	3		91,044,828	605,604	576,959	280,197	102,806
	1	10^6	56,787,933	377,830	358,598	169,092	19,700
	2		77,314,380	514,522	487,870	226,016	27,043
	3		68,665,175	456,572	435,296	210,969	23,631
	1	10^5	67,997,614	451,950	429,353	210,247	3,555
	2		60,372,011	401,234	380,691	181,631	3,168
	3		62,857,231	417,616	398,239	195,931	1,814
	1	N.C.	60,297,197	399,994	386,674	219,579	0
	2		65,545,913	434,856	420,071	212,730	1
	3		67,530,809	447,866	432,325	210,571	41
B4GNG1-2	1	10^8	128,060,503	851,038	816,231	414,953	373,558
	2		109,095,267	725,606	695,131	349,362	267,587
	3		103,081,871	685,992	655,534	323,582	177,227
	1	10^7	115,536,552	768,416	734,298	360,866	159,902
	2		86,686,038	576,230	547,995	274,234	90,922
	3		96,405,761	640,996	610,292	296,217	118,135
	1	10^6	62,068,027	413,482	391,195	185,051	26,659
	2		87,005,474	578,864	550,731	263,528	55,599
	3		71,146,388	473,260	446,615	206,356	37,194
	1	10^5	63,055,893	419,902	396,220	197,320	2,051
	2		75,126,048	499,996	473,196	222,902	10,371
	3		56,503,643	375,320	356,592	178,211	798
	1	N.C.	71,908,602	477,388	462,875	219,006	0
	2		60,639,953	402,194	387,715	197,508	0
	3		52,633,882	349,186	335,987	234,027	41
B1GNG8-2	1	10^8	120,562,993	801,254	765,999	383,366	240,349
	2		132,374,674	880,360	842,857	419,542	253,865
	3		101,512,881	675,110	645,270	322,575	195,132
	1	10^7	99,860,057	663,998	637,040	320,456	122,916
	2		91,555,312	608,690	583,612	292,394	110,905
	3		123,499,309	820,032	787,067	394,779	193,057
	1	10^6	69,778,141	463,358	441,580	223,764	29,109
	2		73,660,615	489,212	469,644	237,799	32,919
	3		66,238,543	440,072	420,973	214,028	16,037
	1	10^5	72,560,042	480,974	461,927	236,193	309

	2		68,935,920	457,018	436,178	223,234	1,879
	3		73,524,495	488,064	467,809	239,006	1,574
	1	N.C.	62,653,766	416,168	401,228	185,585	0
	2		59,709,934	396,218	381,320	193,092	0
	3		58,326,877	387,474	374,416	199,921	0
B1GNS10-2	1	10^8	112,438,475	748,812	718,785	360,441	235,971
	2		121,976,971	811,692	777,439	388,050	230,360
	3		101,291,935	674,346	645,012	321,367	187,065
	1	10^7	104,488,255	695,116	666,547	332,557	163,697
	2		92,995,760	618,334	592,178	295,550	88,708
	3		115,184,019	766,150	733,282	363,411	172,234
	1	10^6	77,869,355	518,622	492,483	236,345	28,517
	2		71,894,289	478,292	459,399	228,707	16,467
	3		76,260,756	507,146	485,085	238,776	34,297
	1	10^5	76,824,446	511,034	484,205	232,124	5,208
	2		80,309,982	535,220	508,834	240,929	3,845
	3		70,496,175	468,728	447,213	218,982	2,538
	1	N.C.	63,724,702	423,400	409,258	176,867	1
	2		62,410,362	414,276	400,895	201,315	0
	3		52,801,398	350,818	338,890	204,450	0

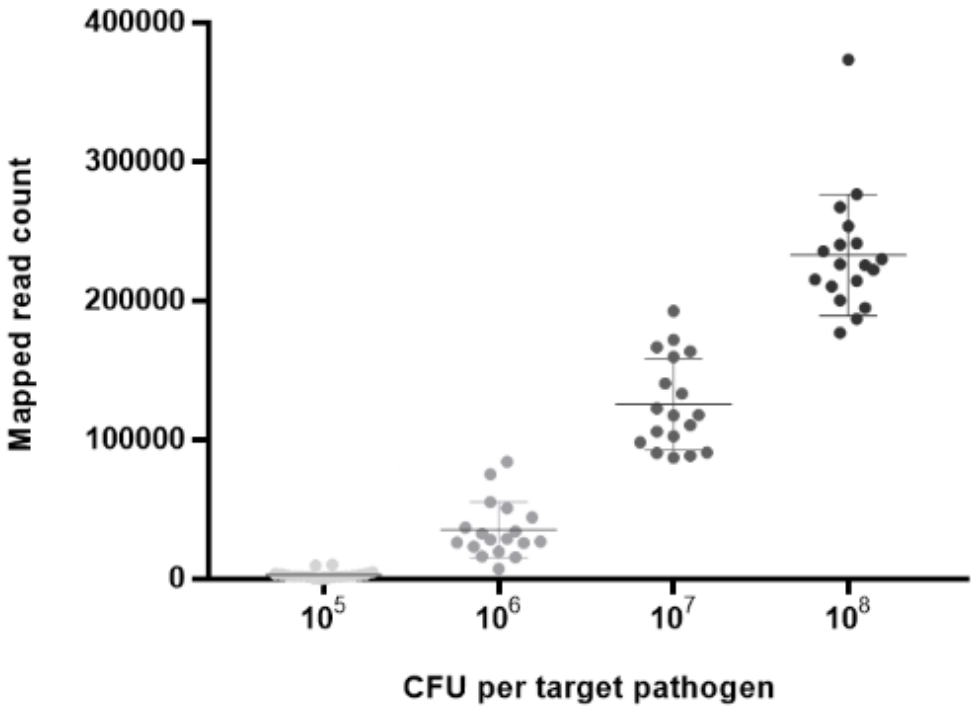

Fig. 5. Dot plot of NGS panel analysis in six agricultural water samples with a mixture of target pathogens.

Contaminated samples with 10^8, 10^7, 10^6, and 10^5 CFU per pathogen were plotted as dot using NGS panel analysis. Each point represents the mean of total target pathogen-specific gene reads in six agricultural water samples.

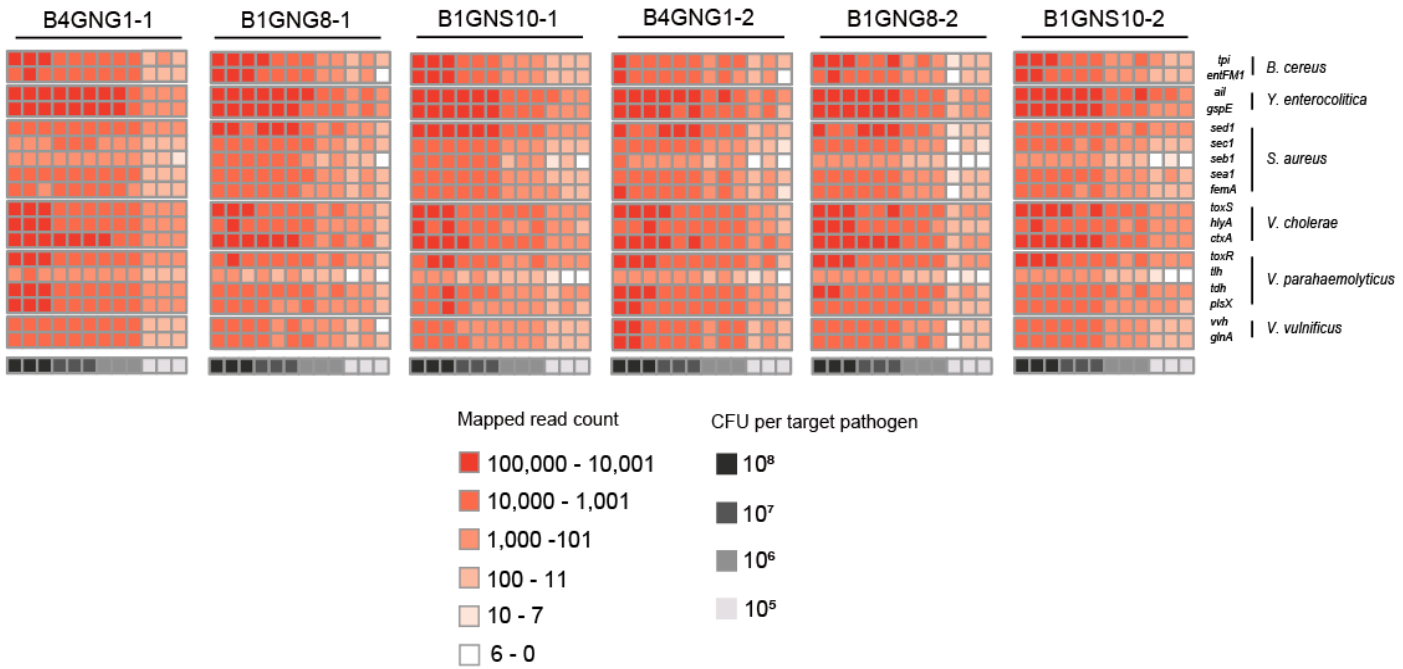

Fig. 6. Heat map of NGS panel analysis result in six agricultural water samples with a mixture of target pathogens. Target pathogen-specific gene reads of the NGS panel were visualized using heat map. The color-scale of target pathogen-specific gene read or target pathogen Ct value and the level of CFU per target pathogen were indicated in the figure.

A

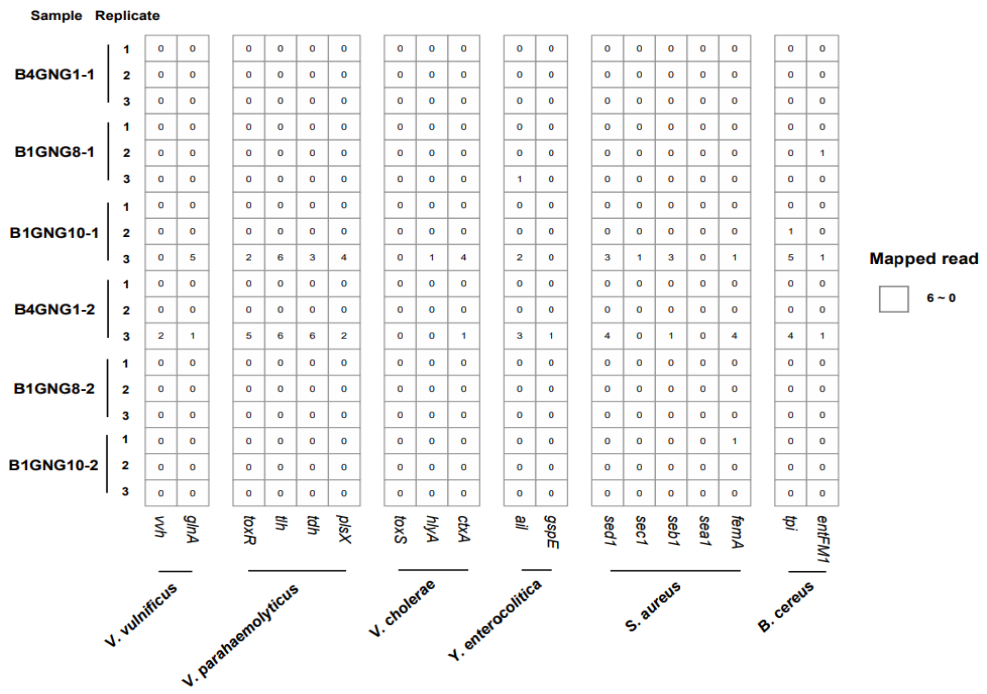

Fig. 7. Target pathogens-specific genes mapped read in six agricultural water samples with or without target pathogens using NGS panel.

(A) Six agricultural water samples without target pathogen.

B

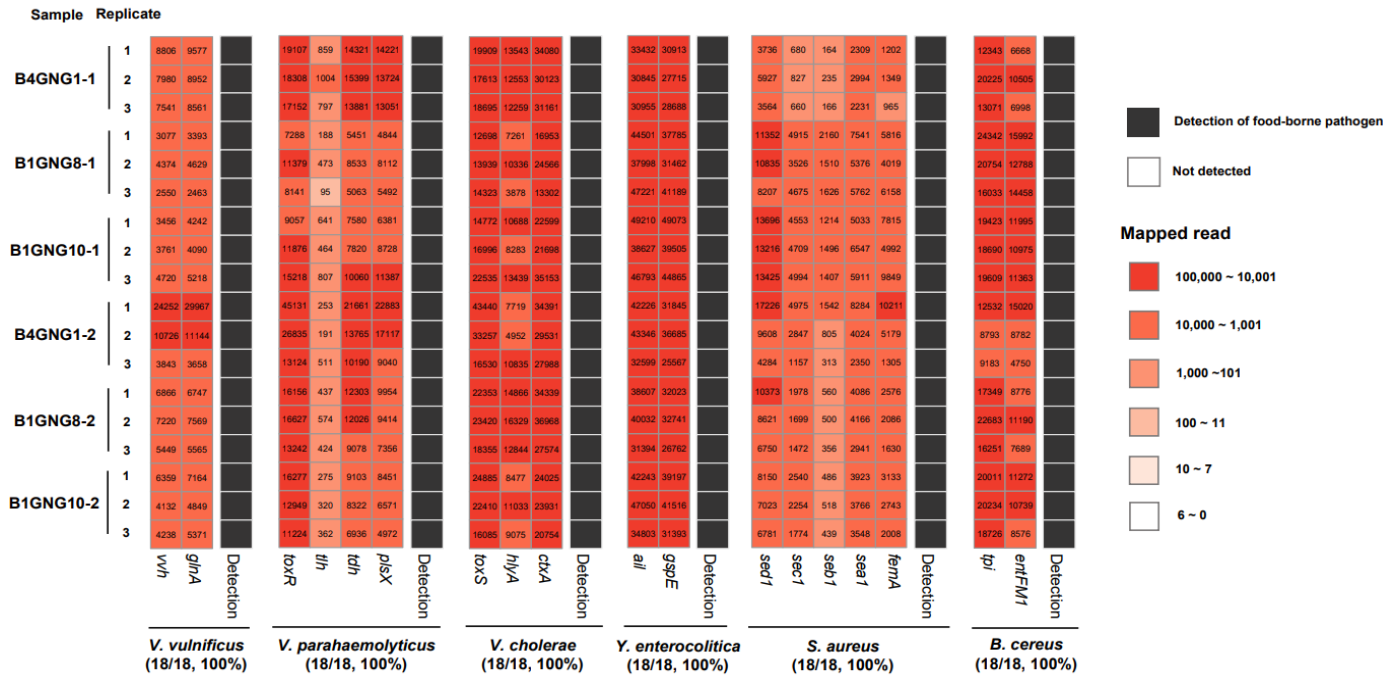

Fig. 7. Cont. Target pathogens-specific genes mapped read in six agricultural water samples with or without target pathogens using NGS panel.

(B) Six agricultural water samples with 10^8 CFU per target pathogen.

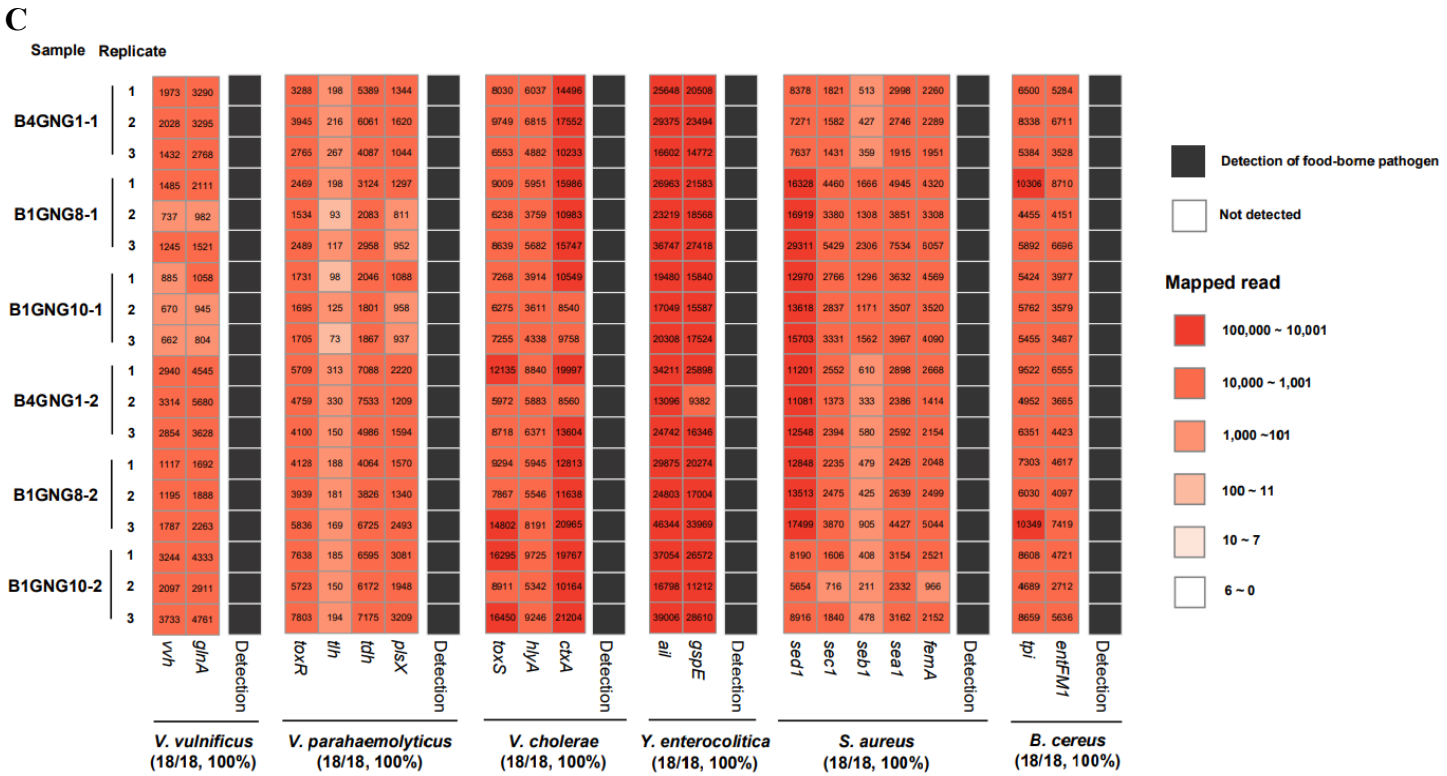

Fig. 7. Cont. Target pathogens-specific genes mapped read in six agricultural water samples with or without target pathogens using NGS panel.

(C) Six agricultural water samples with 10^7 CFU per target pathogen

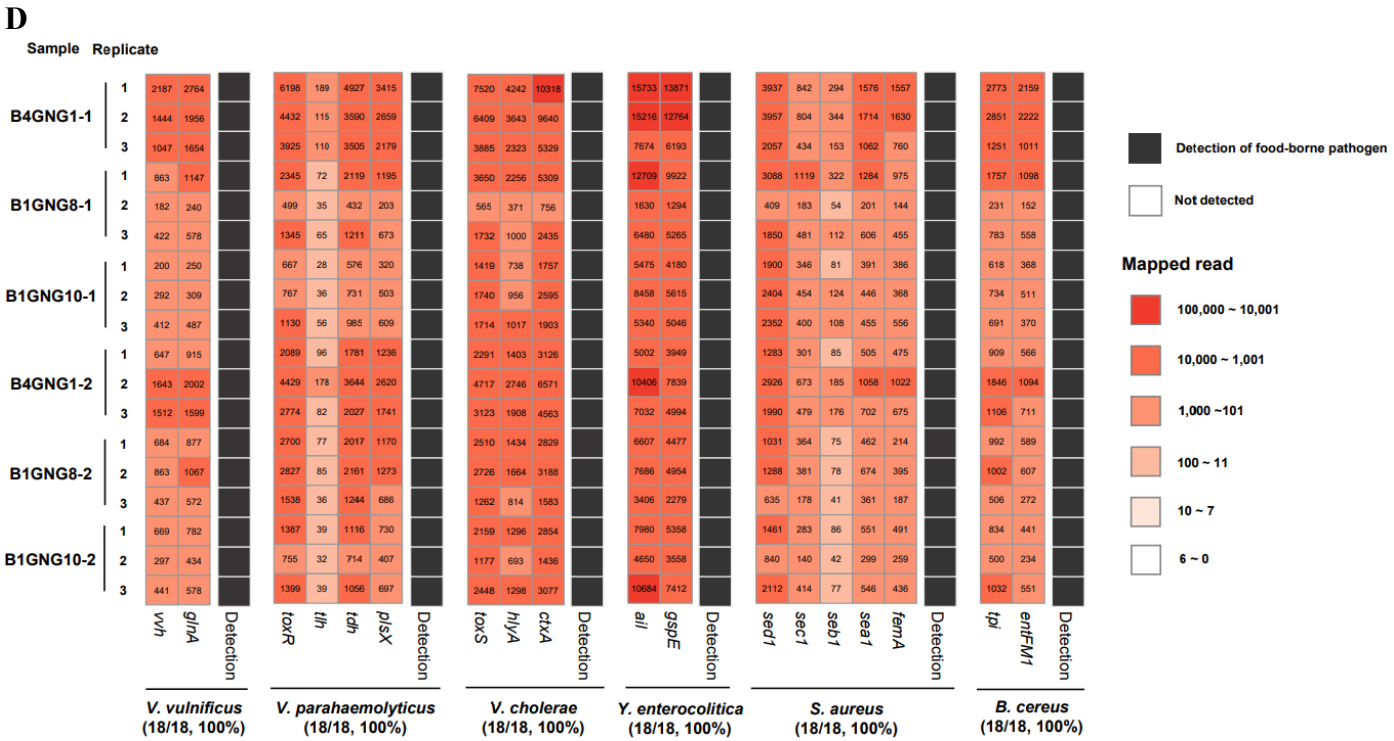

Fig. 7. Cont. Target pathogens-specific genes mapped read in six agricultural water samples with or without target pathogens using NGS panel.

(D) Six agricultural water samples with 10^6 CFU per target pathogen.

E

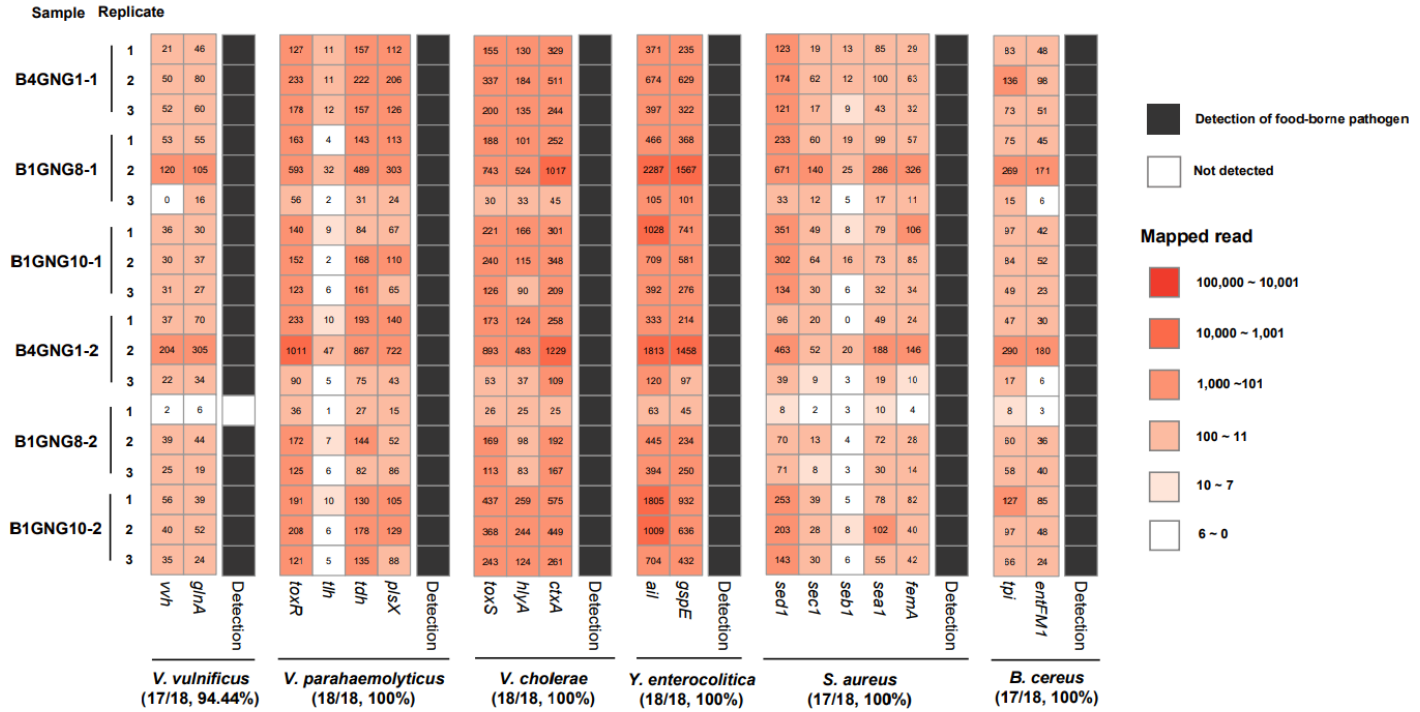

Fig. 7. Cont. Target pathogens-specific genes mapped read in six agricultural water samples with or without target pathogens using NGS panel.

(E) Six agricultural water samples with 10^5 CFU per target pathogen.

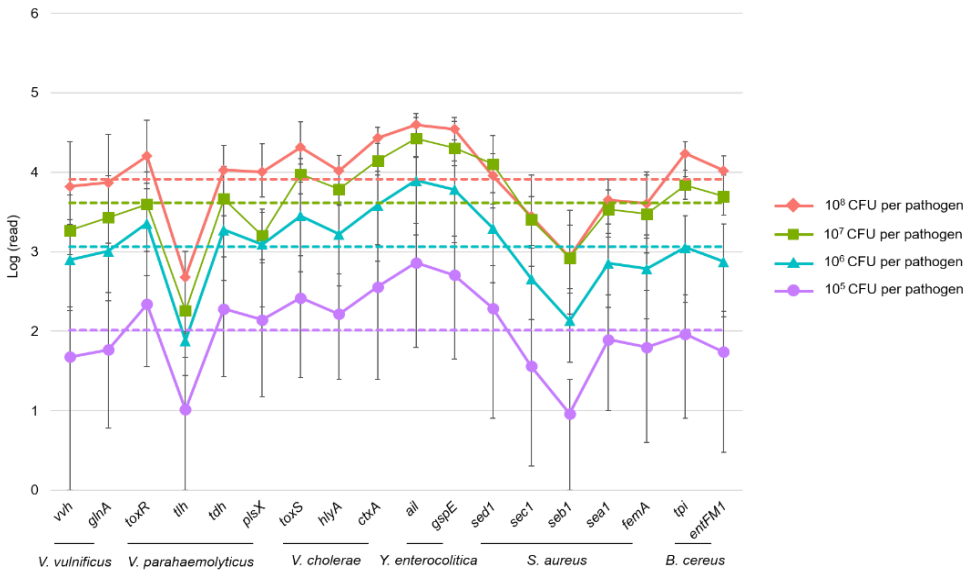

Fig. 8. NGS panel analysis average read summary of target pathogens-specific genes in six agricultural water samples with a mixture of target pathogens.

Mapped read counts to target pathogens-specific genes in 10^8, 10^7, 10^6, and 10^5 CFU per target pathogens were indicated in figure as red, green, blue, and purple, respectively, and dashed line indicates total average read in each CFU per target pathogens. Upper error bar means maximum read counts and lower error bar means minimum read counts.

3.1.5. qPCR analysis

qPCR was conducted for comparison with the NGS panel analysis results. The qPCR template DNA used was the same as that used in the NGS panel analysis. According to the qPCR results, the average Ct values of the target pathogens were 18.84 (10^8 CFUs per target pathogen), 22.33 (10^7), 24.97 (10^6), and 29.49 (10^5)(Fig. 9). Interestingly, a negative correlation existed between Ct and the number of cells of the target pathogen ($y = -3.4584x + 32.556$; $R^2 = 0.9895$), suggesting that target pathogens with a low Ct value or high number of cells could be rapidly detected and identified (Fig. 10). In contrast, the prepared negative control samples lacking specific pathogen contamination showed no Ct values across all qPCR reactions (up to 40 cycles), suggesting that target pathogens were not present in these negative control samples (Fig. 11A).

Furthermore, the Ct values per target pathogen at four different dilution factors (10^8, 10^7, 10^6, and 10^5 CFUs per target pathogen) were compared to determine the sensitivity and detection limit of qPCR (Fig. 10). For all dilution factors, all target pathogens were detected in the qPCR reactions, and the six different target pathogens were identified without false positives (Figs. 11B-E). This result was confirmed in triplicate tests of all agricultural water samples. Overall, these results suggest that the sensitivity of qPCR may be higher than that of NGS panel analysis.

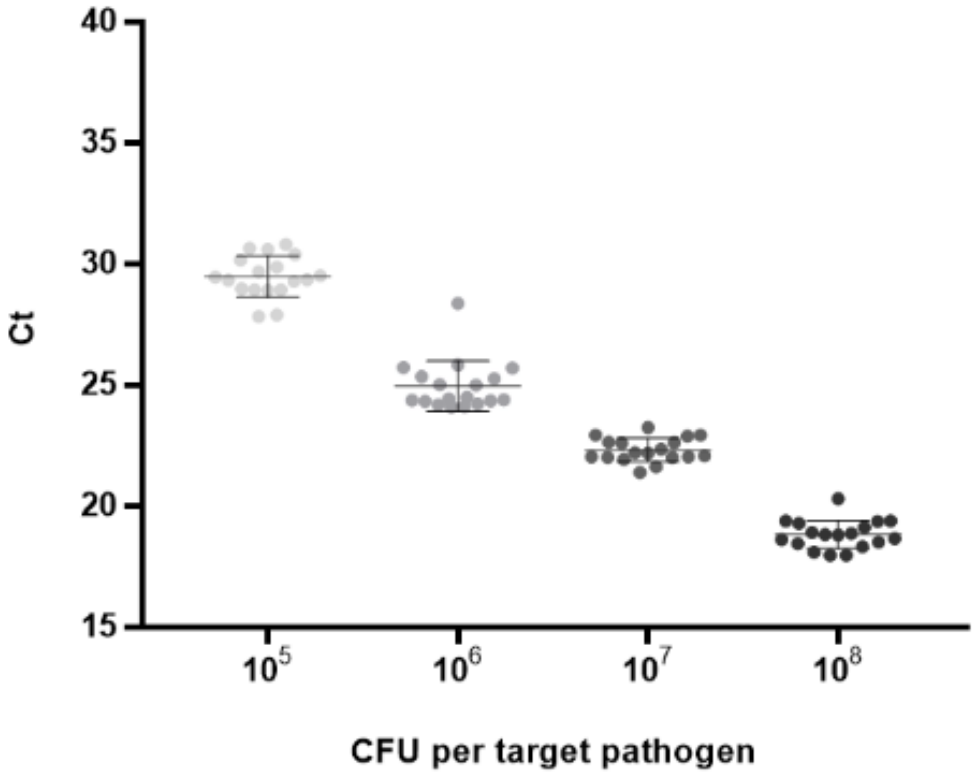

Fig. 9. Dot plot of qPCR results in six agricultural water samples with a mixture of target pathogens.

Ct values of contaminated samples with 10^8, 10^7, 10^6, and 10^5 CFU per pathogen were plotted as dot. Each point represents the mean of target pathogen-specific Ct value.

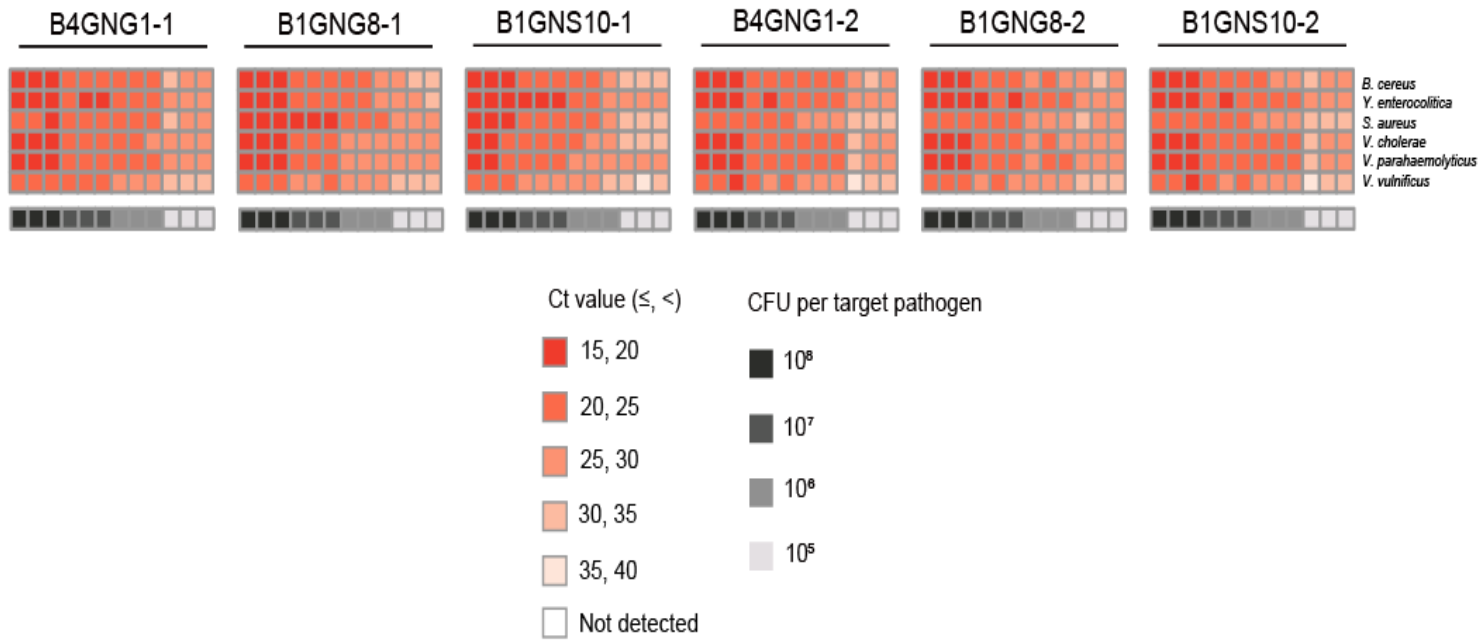

Fig. 10. Heat map of qPCR result in six agricultural water samples with a mixture of target pathogens.

Target pathogen-specific gene reads of the qPCR were visualized using heat map. The color-scale of target pathogen-specific gene read or target pathogen Ct value and the level of CFU per target pathogen were indicated below the figure.

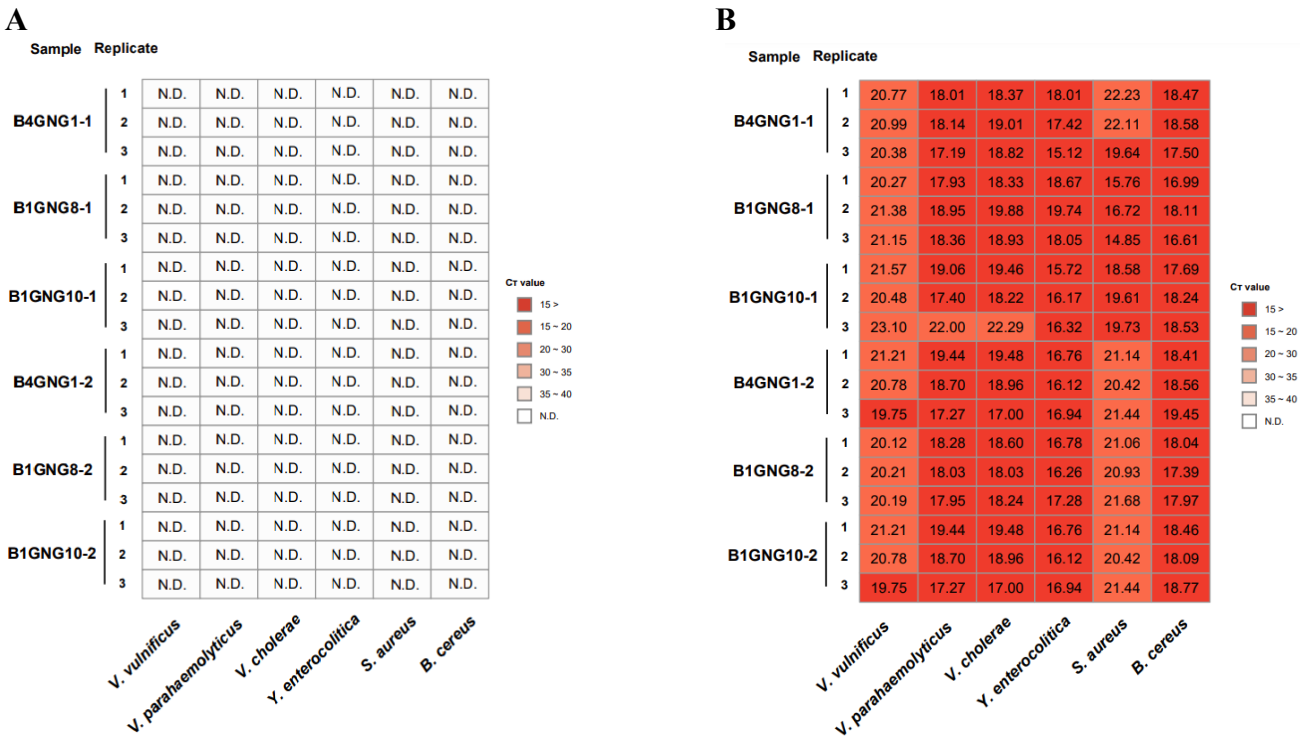

Fig. 11. Ct values of target pathogens in six agricultural water samples with or without target pathogens using qPCR. (A) Six agricultural water samples without target pathogen (B) six agricultural water samples with 10^8 CFU per target pathogen. N.D. indicates Ct values were not exceeded threshold until 40 cycles.

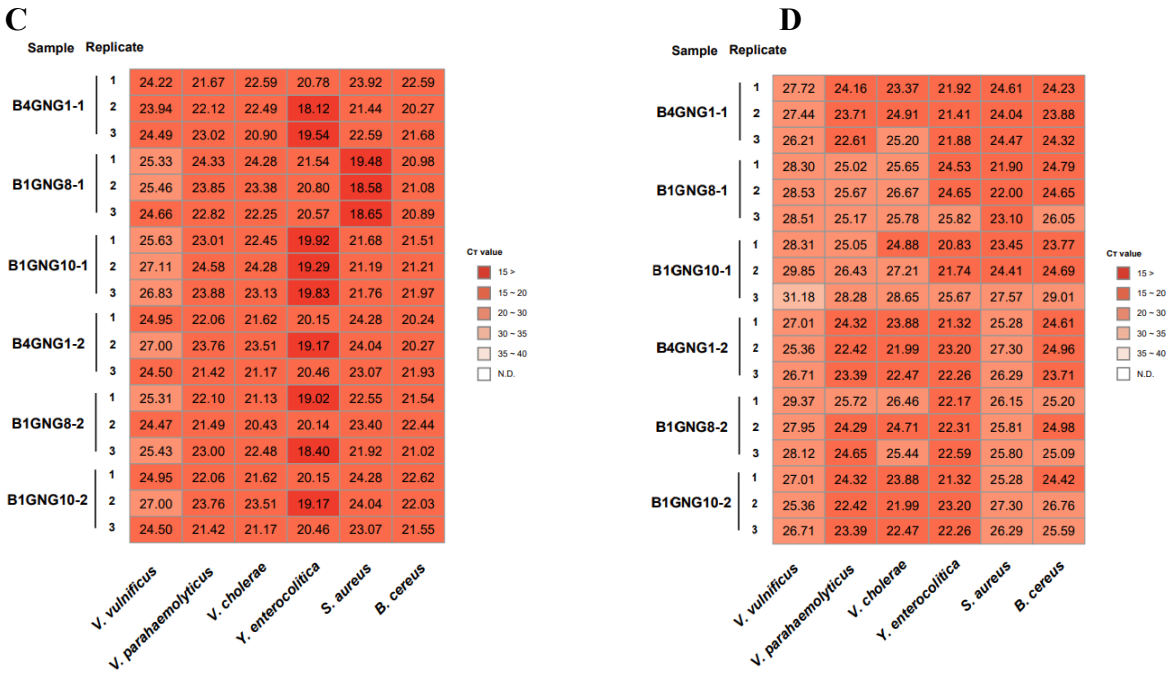

Fig. 11. Cont. Ct values of target pathogens in six agricultural water samples with or without target pathogens using qPCR.

(C) Six agricultural water samples with 10^7 CFU per target pathogen, (D) Six agricultural water samples with 10^6 CFU per target pathogen.

E

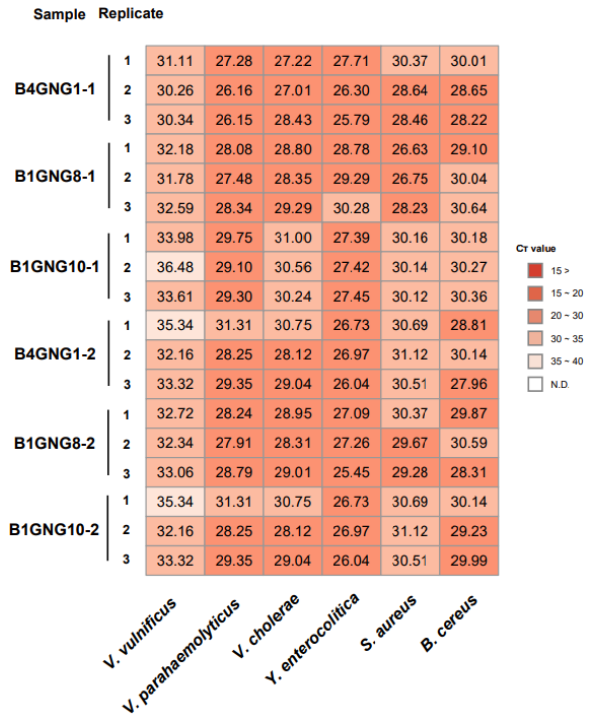

Fig. 11. Cont. Ct values of target pathogens in six agricultural water samples with or without target pathogens using qPCR.

(E) Six agricultural water samples with 10^5 CFU per target pathogen.

3.1.6. Comparative evaluation of NGS panel analysis and qPCR

To further evaluate the detection and identification of specific target pathogens using NGS panel analysis, the results of NGS panel analysis and qPCR were compared. As the qualified read counts and Ct values were correlated with the number of cells of the target pathogens, additional correlation analyses between the qualified read counts and Ct values in each specific target pathogen were performed. Interestingly, the read counts and Ct values were negatively correlated, with comparative analysis revealing a negative relationship ($y = -21154x + 605174$; $R^2 = 0.7984$; Fig. 13), supporting the previous finding that a high number of target pathogen cells is associated with the quicker detection and identification of pathogens. To verify this correlation, Spearman correlation analysis was performed using the results of NGS panel analysis and qPCR for specific target genes, and negative correlations were found in all cases (Fig. 12). The genes *entFM1* and *tpi* of *B. cereus* exhibited the highest correlations, whereas *seb1* of *S. aureus* exhibited the lowest correlation (Fig. 13), which might have been due to the false negative results for this gene in NGS panel analysis (Fig. 7E).

The strong correlation between NGS panel analysis and qPCR in the specific target genes of pathogens suggests that the newly developed NGS panel analysis could serve as a supporting or alternative method to qPCR for the detection and identification of multiple target pathogens in given

environments.

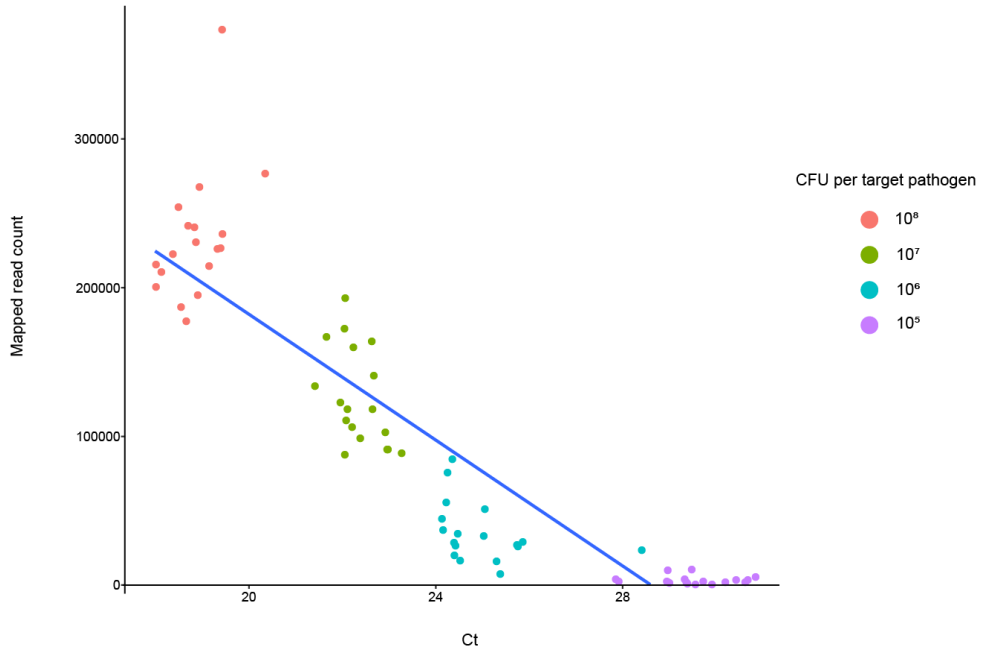

Fig. 12. Linear regression of NGS panel analysis and qPCR results in six agricultural water samples with a mixture of target pathogens.

Contaminated samples with 10^8, 10^7, 10^6, and 10^5 CFUs per pathogen were analyzed using NGS panel analysis and qPCR. Each point is the mean of mapped reads to target pathogen-specific gene or Ct values of each target pathogen in a single replicate. Correlation between the average reads mapped to the total target pathogen-specific genes and average total target pathogen Ct values were described in figure.

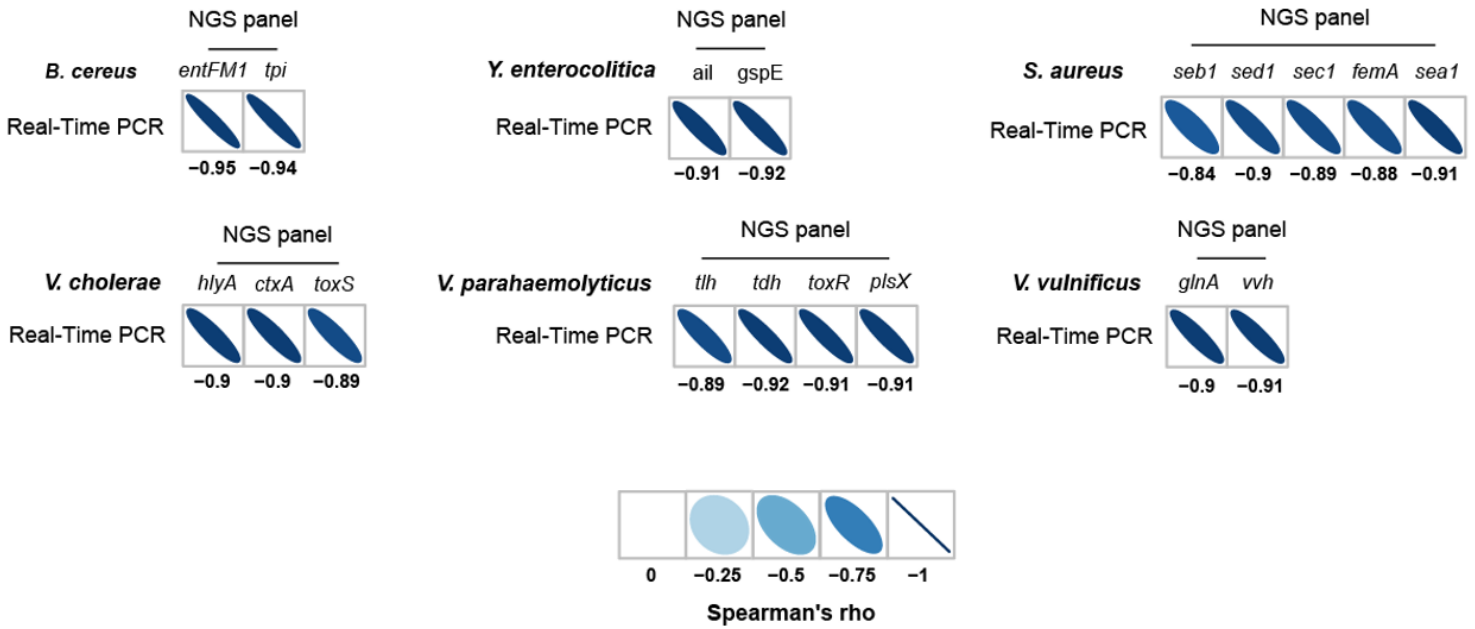

Fig. 13. Spearman correlation between NGS panel analysis and multiplex qPCR in six agricultural water with a mixture of target pathogens.

3.2 NGS panel set 2: multiple detection and identification of foodborne pathogens in fermented foods

3.2.1. Isolation and identification of foodborne pathogens

A total of 88 pathogenic bacteria were isolated from four samples (chicken breast and three animal byproducts). These pathogens included *E. coli* (67 strains), *Listeria monocytogenes* (five strains), *Listeria amylovorus* (one strain), *Salmonella enterica* serovar *Typhimurium* (seven strains), *Streptococcus alactolyticus* (one strain), *Enterococcus faecium* (five strains), and *Bacillus licheniformis* (two strains) in molecular level using 16S rRNA gene sequencing. For further identification of pathogenic types of *E. coli*, pathogen type-specific gene PCR showed the exact PCR amplicon size of *stx1A* (EHEC target gene), *invE* (EIEC target gene), *bfpB* (EPEC target gene), and *elt* (ETEC target gene) gene in four strains of isolated *E. coli* (Fig. 14). Among isolated and identified foodborne pathogens, seven strains, including EHEC SG_006, EIEC SG_007, EPEC SG_010, ETEC SG_009, *L. monocytogenes* SG_004, and *S.* Typhimurium SG_011, were selected as target pathogens. In addition, one EAEC NCCP 14039) was selected (Table 1). Based on previous reports, these foodborne pathogens caused foodborne outbreaks in areas where the fermented food samples were collected (De Buyser, Dufour et al. 2001; Lee, Ha et al. 2018; Oh and Yoon 2017).

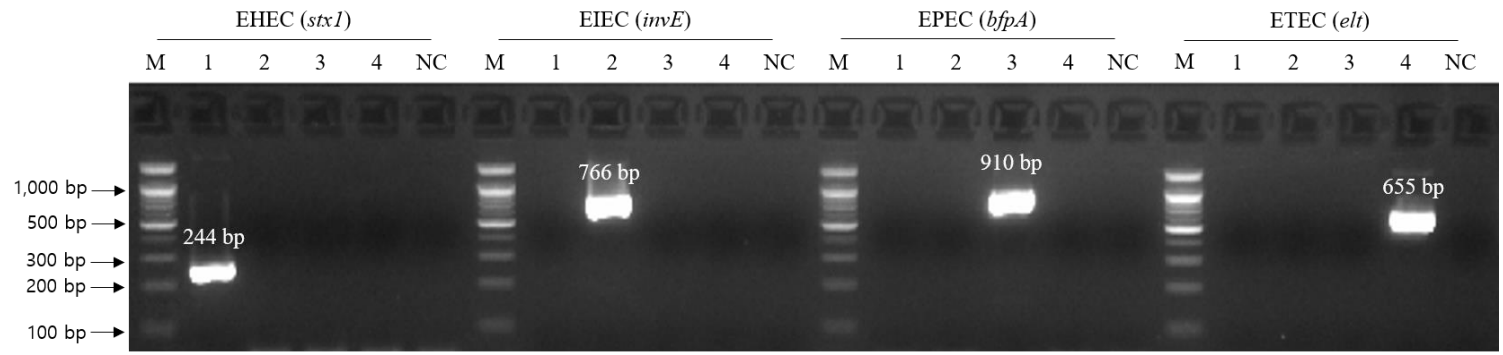

Fig. 14. Gel electrophoresis result of pathogenic type-specific PCR of *E. coli*.

Primer sets targeting each pathogenic type of *E. coli* were described above the Figure. Each lane contains single genomic DNA of isolated *E. coli* strains. Lane 1: EHEC SG_006, lane 2: EIEC SG_007, lane 3: EPEC SG_010, and lane 4: ETEC SG_009. M: 100 bp DNA ladder.

3.2.2 General genome features of selected foodborne pathogens and the design of primer sets

The genome sequence information of selected target pathogens is required to design specific primer sets and confirm their binding sites in the genomes. Therefore, NGS genome sequencing was performed, and draft genome sequences were obtained for *B. cereus*, *Y. enterocolitica*, *V. cholera, V. parahaemolyticus,* and *V. vulnificus* as well as two *S. aureus* strains (SG_001 and CCARM 3089). In addition, the previously reported genome sequences of two *S. aureus* strains (ATCC 23235 and Newman) were obtained from the NCBI GenBank database. The general genome features of these foodborne pathogens are summarized in Table 6. Based on the genome sequences, primer sets targeting specific toxin genes and virulence factors were designed to meet the criteria of primer design given in the Materials and Methods. The sequence information of the designed primer sets is shown in Table 7, and the primer target genes and primer binding locations are listed in Table 8.

Table 6. General genome features of selected foodborne pathogens isolated from animal samples

Bacterium	Strain	Genome size (bp)	Assembly	Contig	GC (%)	CDS	tRNA	rRNA	References
Escherichia coli									
EAEC	NCCP 14039[a]	4,966,374	Draft	105	50.61	4,828	80	4	This study
EHEC	SG_006	5,167,775	Draft	255	50.45	4,889	82	4	This study
EIEC	SG_007	4,927,911	Draft	385	50.78	4,667	51	2	This study
EPEC	SG_010	5,043,792	Draft	330	50.52	4,742	48	3	This study
ETEC	SG_009	5,030,956	Draft	201	50.28	4,794	85	3	This study
Listeria monocytogenes	SG_004	2,962,785	Draft	30	37.93	2,896	47	3	This study
Salmonella enterica serovar Typhimurium	SG_011	4,874,085	Draft	85	52.18	4,571	77	3	This study

[a], NCCP: National Culture Collection for Pathogens

Table 7. Selected pathogen species-specific genes, their functions, and the associated designed primer sets in NGS panel set 2

Bacterium	Gene	Function	Primer	Sequence (5′ to 3′)	Size (bp)	Reference
Escherichia coli						
EAEC	*aggR*	Transcriptional regulator	aggR_F	GATGCTGACGATTCTGTATTA	187	This study
			aggR_R	ATAAGTCCTTCTCGATTGTGT		
EHEC	*stx2A*	Shiga toxin 2 subunit A	stx2A_F	ACTGTCTGAAACTGCTCCTGT	231	This study
			stx2A _R	GGTTGACTCTCTTCATTCACG		
	stxA	Shiga toxin subunit A	stxA_F	GATAGATCCAGAGGAAGGGCG	209	This study
			stxA_R	TACGACTGATCCCTGCAACAC		
EIEC	*invE*	Invasion protein	invE_F	ACGAGTCAACTTTTAGCGAAGGG	234	This study
			invE_R	CTCTATTTCCAATCGCGTCAGAAC		
	stp	Type III secretion system export apparatus protein	stp_F	TCCTGCTTAGATGATGGAGG	173	This study
			stp_R	CCAAAAGGAAGTGTCTGCTC		
EPEC	*bfpA*	Bundle-forming pilus major subunit	bfpA_F	TAGTGGATTGGACTCAACGAT	233	This study
			bfpA_R	TATTAACACCGTAGCCTTTCG		
ETEC	*estB*	Heat-stable enterotoxin ST-I group b	estB_F	CTCAGGATGCTAAACCAGTAGAG	154	This study
			estB_R	CCGGTACAAGCAGGATTACAAC		
	eltA	Heat-labile enterotoxin LT subunit A	eltA_F	TGACGGATATGTTTCCACTTC	191	This study
			eltA_R	GTATTCCACCTAACGCAGAAA		
Listeria monocytogenes	*fusA*	GTP-binding protein	fusA_F	TTGATGGTGCTGTTGCGGTTC	200	This study
			fusA_R	TGGGAGTTGGATTGGGTGC		
	iap	Invasion-associated protein p60	iap_F	CTGGTGATACTCTTTGGGGTA	264	This study
			iap_R	AGCCGTTAGATTCGGTTGTTTC		
	tuf	EF-Tu/IF-2/RF-3 family GTPase	tuf_F	GTGACGAAGTAGAAGTTATCG	198	This study
			tuf_R	AGTTAGTGTGTGGAGTAATCG		
Salmonella enterica serovar Typhimurium	*invA*	Invasion protein	invA_F	CGCACTGAATATCGTACTGG	176	This study
			invA_R	CGATAATTTCACCGGCATCG		
	iapB	Lipopolysaccharide assembly protein	iapB_F	GCTGAGTAACCAACAAGATAA	186	This study

	iapB_R	AGTAAACGCTGTTCATAGGTC

Table 8. *In silico* prediction of primer binding sites in selected foodborne pathogens isolated from animal samples

Bacterium	Strain	Gene	Contig	Primer binding site	
				Forward (nt)	Reverse (nt)
Selected food-borne pathogens					
Enteroaggregative *E. coli* (EAEC)	NCCP 14039	*aggR*	Contig 19	3,410-3,430	3,244-3,264
Enterohaemorhaggic *E. coli* (EHEC)	SG_006	*stx2A*	Contig 109	1,104-1,124	1,314-1,334
		stxA	Contig 93	452-472	640-660
Enteroinvasive *E. coli* (EIEC)	SG_007	*invE*	Contig 48	3,150-3,172	3,360-3,383
		stp	Contig 48	30,537-30,556	30,690-30,709
Enteropathogenic *E. coli* (EPEC)	SG_010	*bfpA*	Contig 48	18,505-18,525	18,717-18,737
Enterotoxigenic *E. coli* (ETEC)	SG_009	*estB*	Contig 239	-	153-174
		eltA	Contig 112	2,691-2,711	2,861-2,881
Listeria monocytogenes	SG_004	*fusA*	Contig 2	448,034-448,054	448,215-448,233
		iap	Contig 1	537,576-537,596	-
		tuf	Contig 2	450,653-450,673	450,830-450,850
Salmonella enterica	SG_011	*invA*	Contig 1	417,113-417,132	417,269-417,288
serovar *Typhimurium*		*iapB*	Contig 11	26,185-26,205	26,350-26,370

3.2.3 Validation of designed primer sets

3.2.3.1. Singleplex PCR

To evaluate the specificity of primer sets to the target pathogens, singleplex PCR was performed with a single target pathogen and an associated single primer set. For the target pathogens, the selected specific genes with their encoded functions, designed specific primer sets, and expected PCR amplicon sizes are listed in Table 7. Following singleplex PCR, agarose gel electrophoresis analysis revealed that all PCR amplicons were of the expected size according to the single PCR bands, confirming the specificity of all the PCR primer sets to the associated target pathogens (Fig. 15). Thus, the designed primer sets qualified for crosscheck PCR evaluation in the next stage.

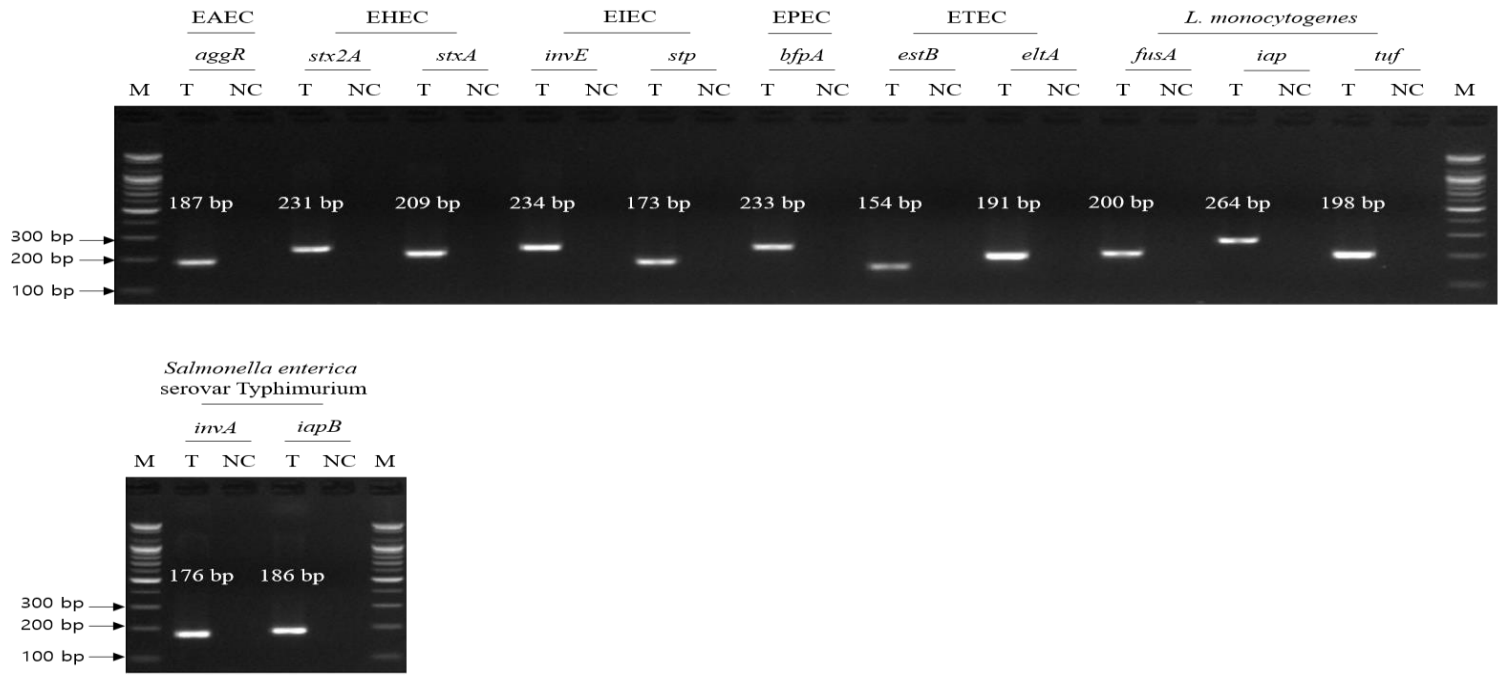

Fig. 15. Gel electrophoresis results of singleplex PCR using NGS panel set 2 primer sets.

Target pathogens and their specific genes are shown above the gel electrophoresis results. PCR mixture of the test (T) lane contained the associated target pathogen genomic DNA and specific gene primer set. PCR mixture of the negative control (NC) lane contained molecular water and the target pathogen-specific gene primer set. M: 100 bp DNA ladder.

3.2.3.2. Crosscheck PCR

In evaluating the specificity of the primer, two different cross-check PCRs were conducted: (a) a single primer set with a genomic DNA mixture of the associated target pathogen and six different non-target pathogens; (b) a single selected pathogen-specific gene primer set (two to three primer sets) mixed with an associated target pathogen.

For the first cross-check PCR, two types of genomic DNA templates were used: (a) test DNA template containing genomic DNA of target and non-target pathogens and (b) negative control DNA template containing only genomic DNA of non-target pathogens. Such templates were prepared to confirm the nonspecific binding of a single selected primer set to the genomic DNA of non-target pathogens. The gel electrophoresis result of the first cross-check PCR showed that the selected target gene-specific PCR amplicon bands were only observed in the test lanes, but no PCR amplicon bands were observed in the negative test lanes (Fig. 16). In addition, the sizes of PCR amplicon bands were similar to the expected ones, indicating that such primer sets are specific to the genomic DNA of target pathogens, although the DNA template contains all other genomic DNA of the non-target pathogens. Based on the first cross-check PCR results, primer sets are specific to the associated target gene and target pathogen.

The second cross-check PCR was performed to evaluate whether one

PCR (with multiple primer sets targeting single pathogen-specific genes) can multi-detect the target genes in a single pathogen. In particular, primer sets of EAEC and EPEC were omitted from the second cross-check PCR because only single-target gene was selected. Therefore, the second cross-check PCR primer set is a mixture of the primer sets targeting two to three selected genes in a single pathogen (a total of five combinations of primer mixture, Table 7). The gel electrophoresis result of the second cross-check PCR revealed that the PCR amplicons of all target genes in each pathogen were confirmed in the gel, and their amplicon sizes were similar to the expected ones (Fig. 17). Therefore, PCR with the mixture of primer sets targeting single-target pathogen-specific genes can detect target genes in one reaction without any primer interference.

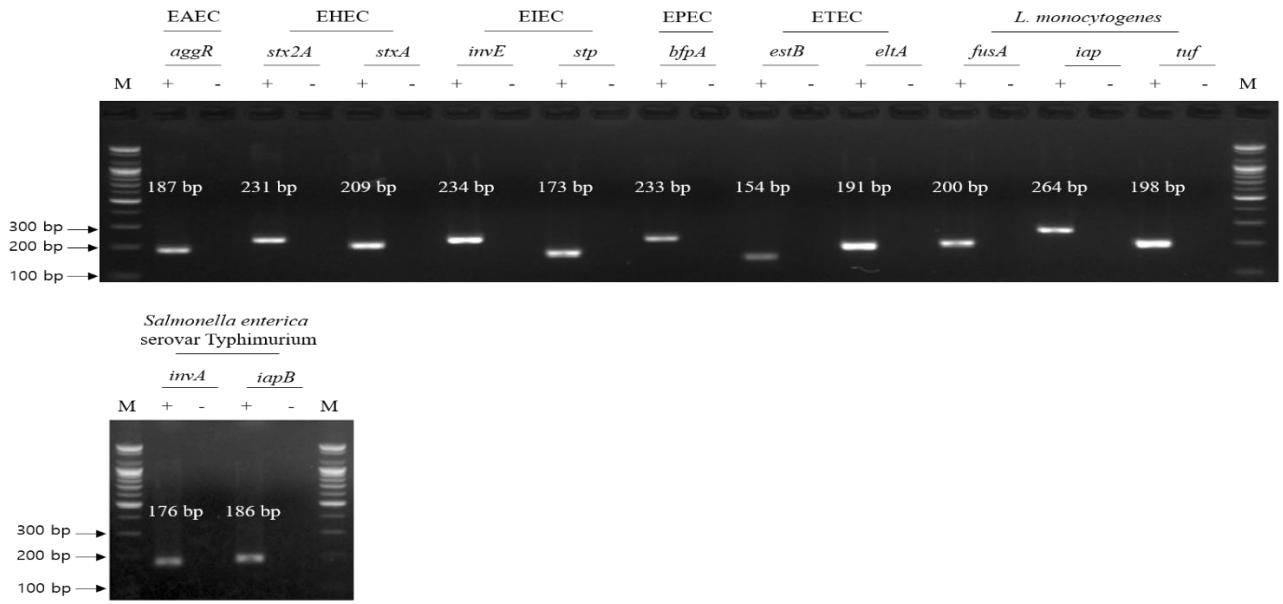

Fig. 16. Gel electrophoresis results of the first crosscheck PCR using NGS panel set 2 primer sets.

Target pathogens and their specific genes are shown above the gel electrophoresis results. PCR mixture of the test (+) lane contained a genomic DNA mixture including the associated target pathogen and target pathogen-specific gene primer set. PCR mixture of the negative test (−) lane contained a genomic DNA mixture lacking the associated target pathogen and target pathogens-specific gene primer set. M: 100 bp DNA ladder.

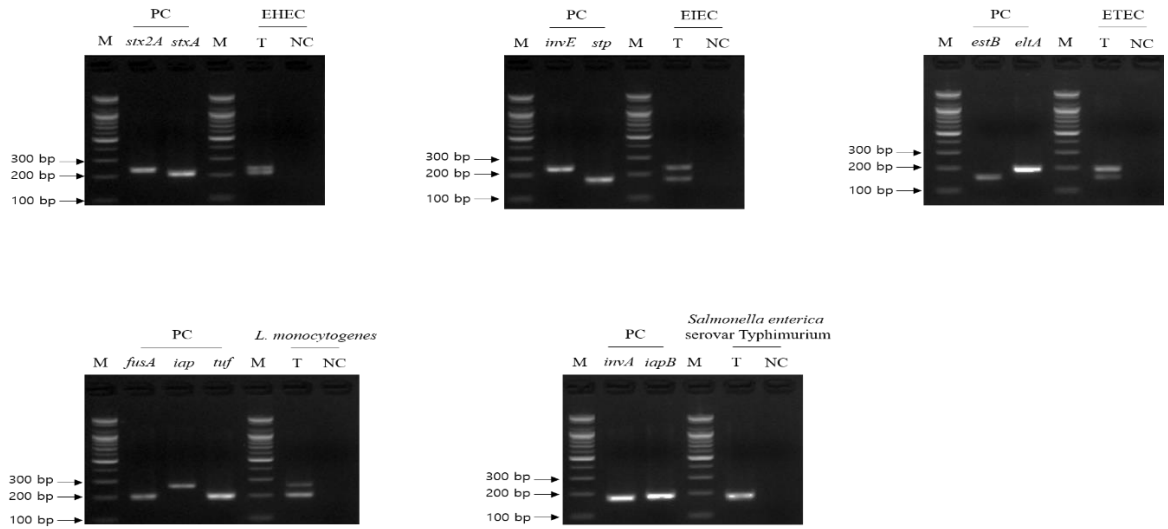

Fig. 17. Gel electrophoresis results of the second crosscheck PCR using NGS panel set 2 primer sets.

Target pathogens and their specific genes are shown above the gel electrophoresis results. Lanes contain each target pathogen-specific gene with singleplex PCR amplicons as positive controls. PCR mixture of the test (T) lane contained the associated target pathogen genomic DNA and two to three target pathogen-specific gene primer sets. PCR mixture of the negative control (NC) lane contained molecular water and two to three target pathogen-specific gene primer sets. M: 100 bp DNA ladder.

3.2.3.3. Multiplex PCR

Based on the results of the first cross-check and second cross-check PCRs, multiple target genes could be detected in one PCR, although primer sets and several pathogenic DNAs were mixed. Hence, multiplex PCR was performed with the mixture of primer set and several pathogenic DNAs. In particular, EAEC and EPEC targeting primer set mixtures were not tested in multiplex PCR because only single-target genes were selected.

For the multiplex PCR, DNA templates were prepared using the same procedure as that of the first cross-check PCR. In addition, the mixture of primer sets was prepared using the same mixture as that of the second cross-check PCR. The gel electrophoresis result of multiplex PCR showed that PCR amplicons of all multiple target genes per selected pathogen were detected in the gel, and their band sizes were the same to the expected ones (Fig. 18). Therefore, these primer sets will be susceptible for further NGS panel analysis.

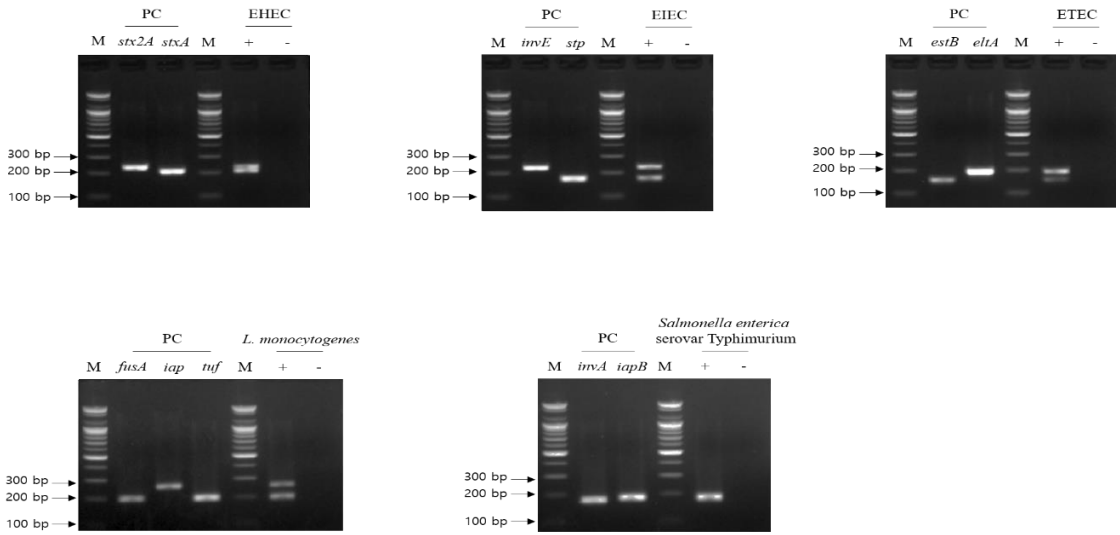

Fig. 18. Gel electrophoresis results of the multiplex PCR using NGS panel set 2 primer sets.

Target pathogens and their specific genes are shown above the gel electrophoresis results. Lanes contain each target pathogen-specific gene with singleplex PCR amplicons as positive controls. PCR mixture of the test (+) lane contained a genomic DNA mixture including the associated target pathogen and two to three target pathogen-specific gene primer sets. PCR mixture of the negative test (−) lane contained a genomic DNA mixture lacking the associated target pathogen and 2−5 target pathogens-specific gene primer sets. M: 100 bp DNA ladder.

3.2.4 NGS panel analysis

NGS panel analysis was performed with six different fermented food samples contaminated with the mixture of seven target pathogens. The NGS panel results showed that the average of the mapped sequence read to target pathogen-specific genes was obtained: 161,081 (54.77% of total qualified sequence reads), 28,929 (14.45%), 1,765 (1.23%), and 237 (0.15%) at 10^8, 10^7, 10^6, and 10^5 CFU per target pathogen, respectively (Table 9 and Fig. 19). In addition, the average of mapped sequence reads to target pathogen-specific genes and CFU per target pathogen was proportional (Fig. 19). However, the prepared negative control without contamination in samples showed 1 to 3 mapped reads to target pathogen-specific genes in NGS panel analysis, indicating that a small number of those pathogens might be present in the original fermented food samples as a false-positive (Fig. 21A). Therefore, ≤3 reads were determined to be a false-positive for further NGS panel analysis.

After mapping to 13 different target genes of seven target pathogens, all qualified NGS panel sequence reads were collected from six different fermented food samples. Then, the collected read counts in each dilution factor (10^8, 10^7, 10^6, and 10^5 CFU per target pathogen) were compared for the detection and identification of specific target pathogens (Fig. 20). In dilution factors of 10^7 to 10^8, all 13 target genes multiplied, which was

sufficient to identify seven different target pathogens in one NGS panel analysis without a false-positive (Figs. 21B and 21C). In addition, this result was confirmed in triplicate tests of all agricultural water samples. The serial dilution of target pathogens was proportionally associated with the read counts, showing the highest number of read counts in dilution factor of 10^8 and lowest number of 10^7, which is consistent with the result shown in Fig. 20. However, in dilution factor of 10^6, false-positive results were only detected in the *fusA* gene of *L. monocytogenes* (Fig. 21D). Furthermore, many false-positive reads appeared in dilution factor of 10^5 (Fig. 21E). In particular, the *stxA* gene of EHEC and the *fusA* gene of *L. monocytogenes* were poorly detected by NGS panel analysis (Figs. 21E and 22). Therefore, these two genes may be removed to increase the limits of detection and identify specific target pathogens in NGS panel analysis. Finally, this result indicates that the limits of detection and identification of target pathogens may be 10^7 CFU.

Table 9. Summary of NGS panel set 2 outputs in six fermented food samples with or without target pathogen

Sample	Replicate	CFU	Yield (bp)	Raw read	Filtered read	Merged read	Mapped read to total target pathogen-specific gene
Cabbage Kimchi	1	10^8	53,429,124	355,190	339,952	165,794	70,542
	2		69,651,085	462,872	441,996	220,140	107,531
	3		45,070,533	299,620	286,467	141,013	59,030
	1	10^7	36,957,058	246,544	236,411	106,954	18,796
	2		31,106,357	207,564	198,680	93,804	6,017
	3		42,170,573	281,004	269,764	126,422	8,999
	1	10^6	35,832,348	238,836	228,900	104,828	1,284
	2		30,345,896	202,458	194,198	87,696	1,252
	3		34,723,694	231,976	222,001	94,260	2,392
	1	10^5	42,269,479	281,864	270,533	126,874	304
	2		33,933,097	226,258	215,239	101,478	212
	3		39,175,130	260,826	249,793	121,741	184
	1	N.C.	38,116,622	253,854	244,994	113,261	0
	2		33,716,443	224,394	216,563	108,260	0
	3		35,405,285	235,984	228,008	122,988	0
Raddish Kimchi	1	10^8	119,029,566	790,314	758,780	375,276	236,167
	2		92,341,378	613,376	588,105	284,809	159,487
	3		79,112,446	525,722	503,011	237,117	121,950
	1	10^7	64,101,041	426,938	408,025	179,054	18,574
	2		60,770,807	404,838	386,337	170,721	26,261
	3		63,346,114	422,010	402,598	173,925	32,547
	1	10^6	44,372,781	295,882	282,331	106,064	1,795
	2		56,277,277	375,088	358,322	129,738	1,877
	3		51,756,096	344,084	329,529	143,460	1,850
	1	10^5	42,264,437	281,760	268,996	116,120	175
	2		54,255,943	361,372	345,183	143,934	159
	3		36,070,553	240,524	229,628	91,684	123
	1	N.C.	45,089,840	301,420	290,931	118,670	1
	2		35,130,671	234,980	226,662	112,608	2
	3		37,241,290	249,118	240,200	143,976	1

Leaf mustard Kimchi	1	10^8	102,131,433	678,362	648,311	329,337	196,990
	2		114,155,954	758,706	725,247	366,860	227,464
	3		69,927,117	464,952	442,727	225,349	128,134
	1	10^7	57,899,344	384,348	369,551	183,601	16,458
	2		62,753,641	416,964	400,619	191,505	35,646
	3		49,583,195	329,482	315,545	154,093	22,435
	1	10^6	54,413,409	361,426	346,535	169,068	2,032
	2		47,695,192	316,630	303,300	149,370	1,553
	3		54,509,836	361,950	348,775	170,186	2,190
	1	10^5	53,041,047	352,716	340,371	162,866	228
	2		59,655,782	396,480	381,374	182,572	333
	3		60,192,138	400,126	384,641	188,017	181
	1	N.C.	40,838,316	272,172	261,966	137,520	1
	2		40,342,578	269,330	259,930	129,251	1
	3		42,880,990	286,462	276,722	130,875	7
Greek yogurt	1	10^8	79,557,479	527,854	504,998	247,209	142,687
	2		78,654,898	521,876	499,592	245,062	114,696
	3		85,604,438	567,800	541,170	272,147	138,466
	1	10^7	45,085,584	299,100	286,388	139,856	12,841
	2		45,038,325	298,690	284,233	142,209	10,128
	3		45,405,151	301,044	289,093	145,045	4,260
	1	10^6	50,406,799	334,258	321,289	161,604	780
	2		38,689,520	256,740	245,803	118,695	758
	3		42,624,682	282,578	269,353	135,503	1,148
	1	10^5	55,252,280	366,442	350,978	175,586	78
	2		40,496,261	268,656	255,599	129,743	326
	3		51,803,697	343,600	329,753	162,844	134
	1	N.C.	34,217,528	226,874	219,818	113,679	3
	2		37,755,625	250,352	242,366	123,292	0
	3		34,796,091	230,728	223,668	111,570	6
Yoghurt	1	10^8	123,533,914	819,892	787,588	401,774	255,198
	2		113,324,200	752,128	723,773	368,635	232,725
	3		108,881,103	722,692	693,206	353,341	229,070
	1	10^7	74,145,666	492,170	472,729	238,512	54,318
	2		86,573,630	574,806	552,963	278,651	101,121
	3		86,356,506	573,240	553,005	280,113	91,830
	1	10^6	57,250,201	380,058	363,335	184,915	4,418
	2		54,762,577	363,484	348,508	177,073	3,157

	3		60,084,440	398,596	380,619	194,364	2,371
	1	10^5	56,318,220	373,744	348,666	178,401	147
	2		58,463,476	388,228	373,786	189,204	991
	3		60,740,183	403,042	384,623	196,179	299
	1	N.C.	36,887,613	244,732	237,096	112,302	1
	2		39,327,825	260,890	252,150	127,852	1
	3		34,618,606	229,710	222,015	119,947	0
Liquid yogurt	1	10^8	97,580,132	647,754	619,965	315,706	182,400
	2		95,286,399	632,424	607,001	309,126	150,742
	3		90,083,525	597,908	572,975	291,904	146,181
	1	10^7	61,736,114	410,022	394,769	200,105	33,175
	2		50,732,505	336,658	322,687	164,636	3,311
	3		58,246,572	386,644	371,913	188,512	24,007
	1	10^6	51,357,594	340,764	327,406	166,586	1,080
	2		51,858,764	343,930	328,281	168,031	542
	3		53,581,987	355,512	340,438	173,389	1,296
	1	10^5	53,275,387	353,464	339,396	173,068	134
	2		59,637,304	395,996	380,513	193,341	193
	3		52,820,435	350,480	337,863	171,627	64
	1	N.C.	39,146,446	260,220	249,681	116,560	2
	2		38,573,027	256,254	245,107	123,915	2
	3		36,655,209	243,618	232,168	126,548	0

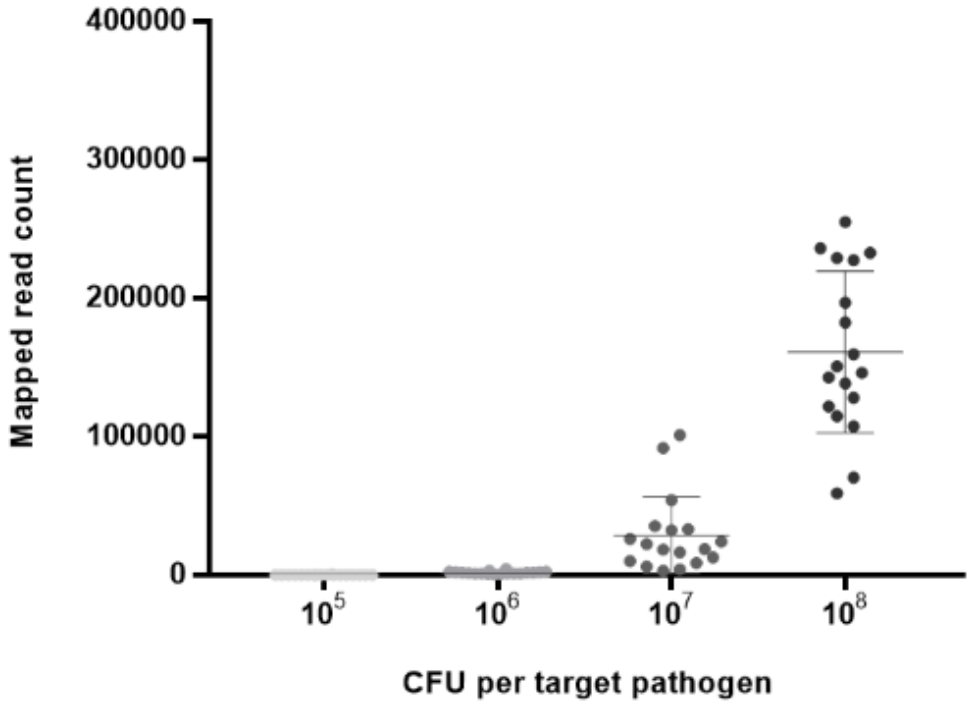

Fig. 19. Dot plot of NGS panel analysis in six fermented food samples with a mixture of target pathogens.

Contaminated samples with 10^8, 10^7, 10^6, and 10^5 CFU per pathogen were plotted as dot using NGS panel analysis. Each point represents the mean of total target pathogen-specific gene reads in six fermented food samples.

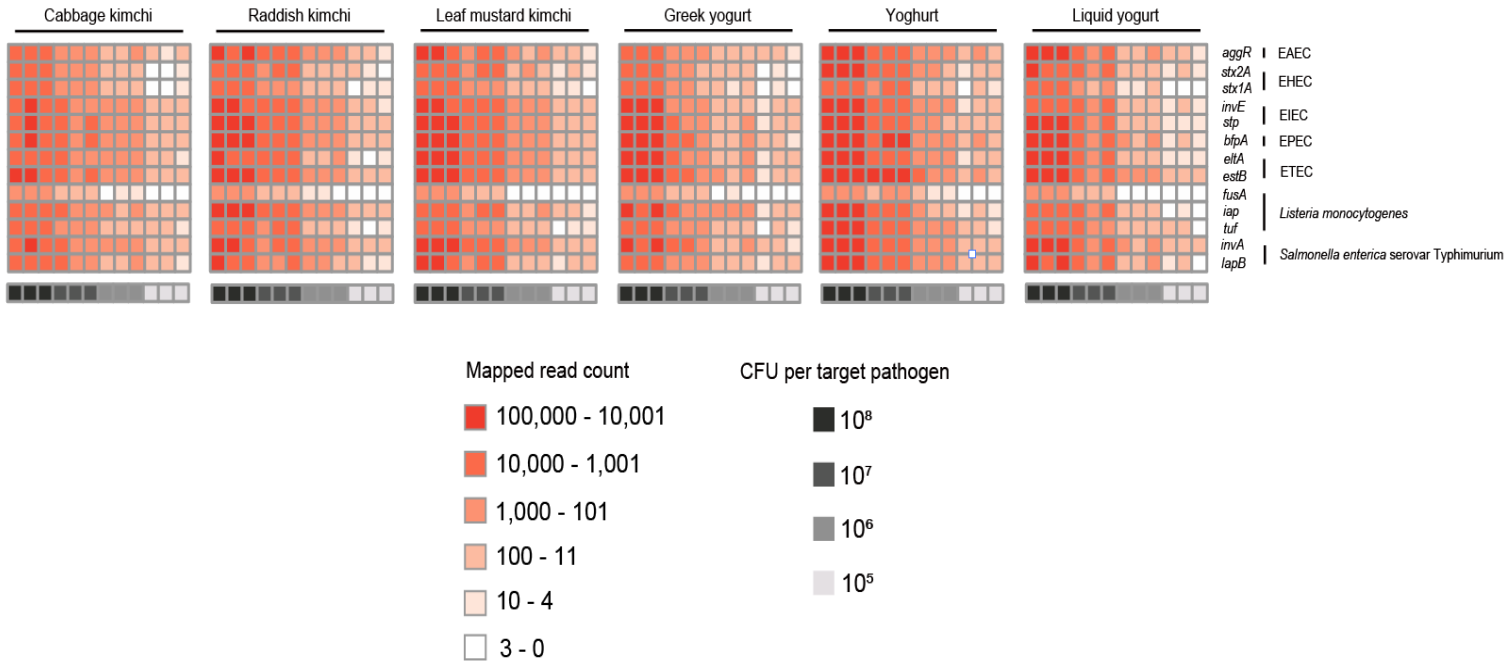

Fig. 20. Heat map of NGS panel analysis result in six fermented food samples with a mixture of target pathogens. Target pathogen-specific gene reads of the NGS panel were visualized using heat map. The color-scale of target pathogen-specific gene read or target pathogen Ct value and the level of CFU per target pathogen were indicated in the figure.

A

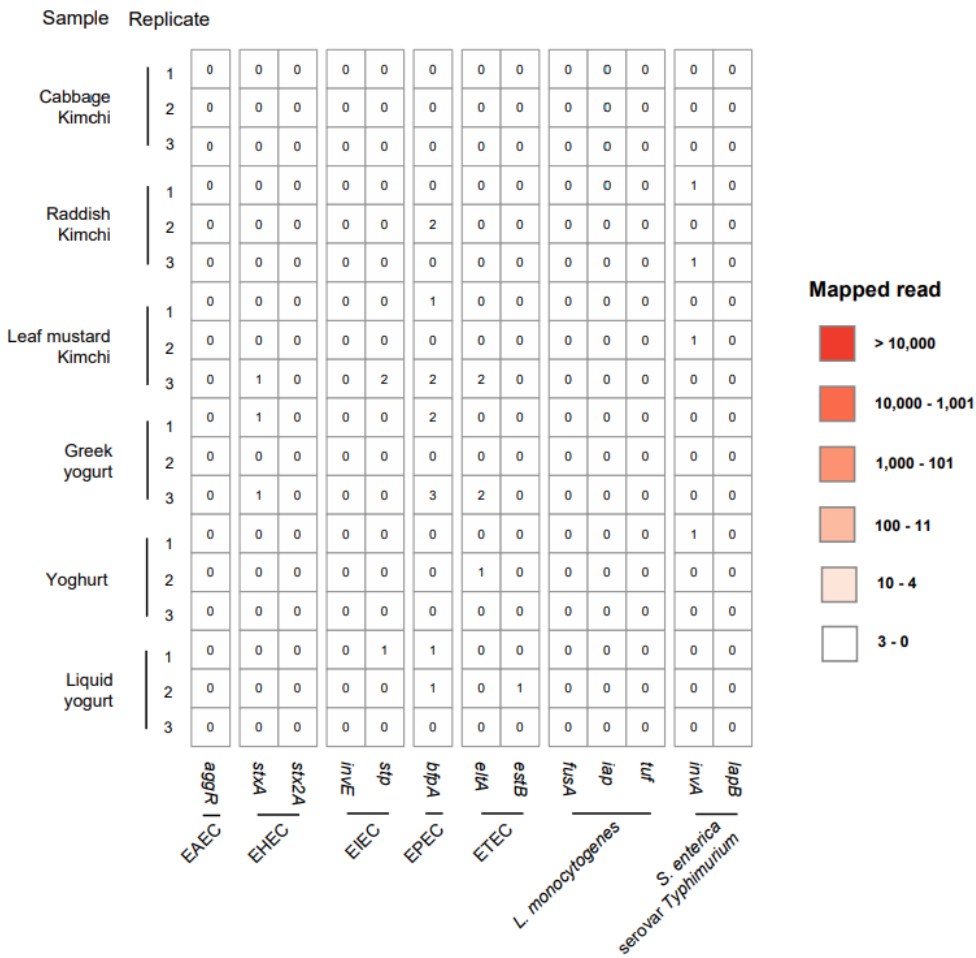

Fig. 21. Target pathogens-specific genes mapped read in six fermented food samples with or without a mixture of target pathogens.

(A) Six fermented food samples without target pathogen.

B

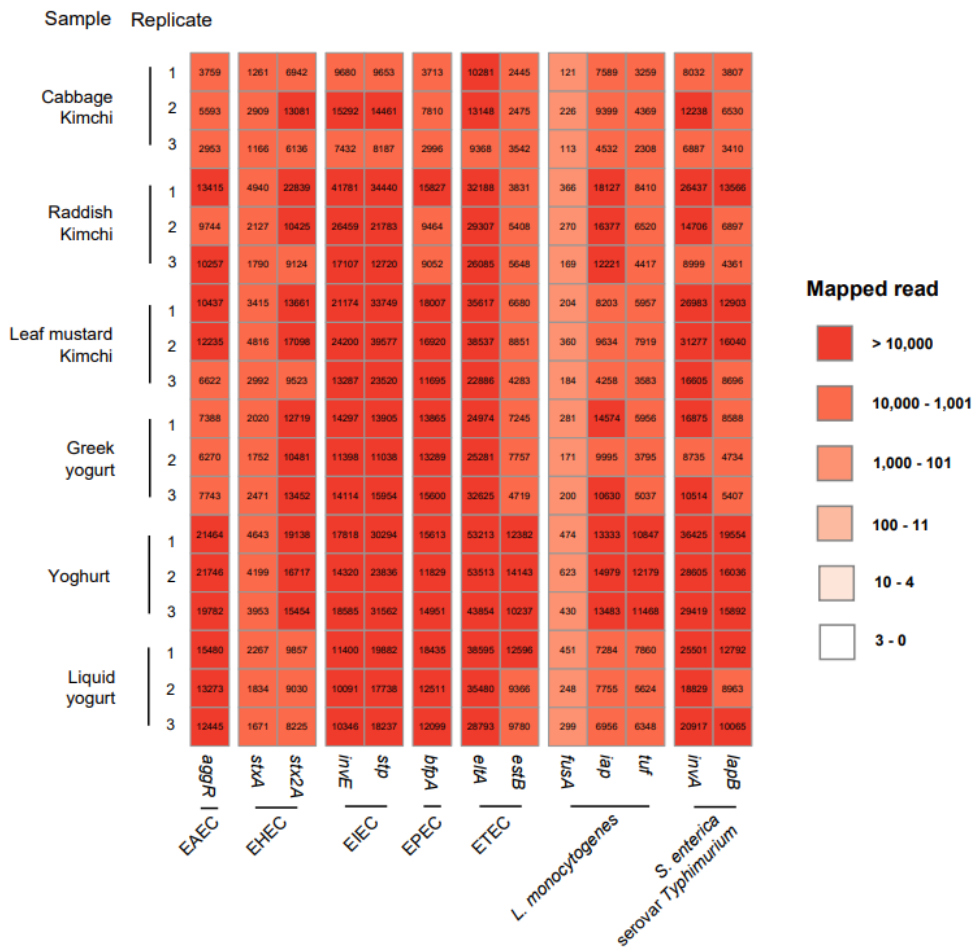

Fig. 21. Cont. Target pathogens-specific genes mapped read in six fermented food samples with or without a mixture of target pathogens.

(B) Six fermented food samples with 10^8 CFU per target pathogen.

C

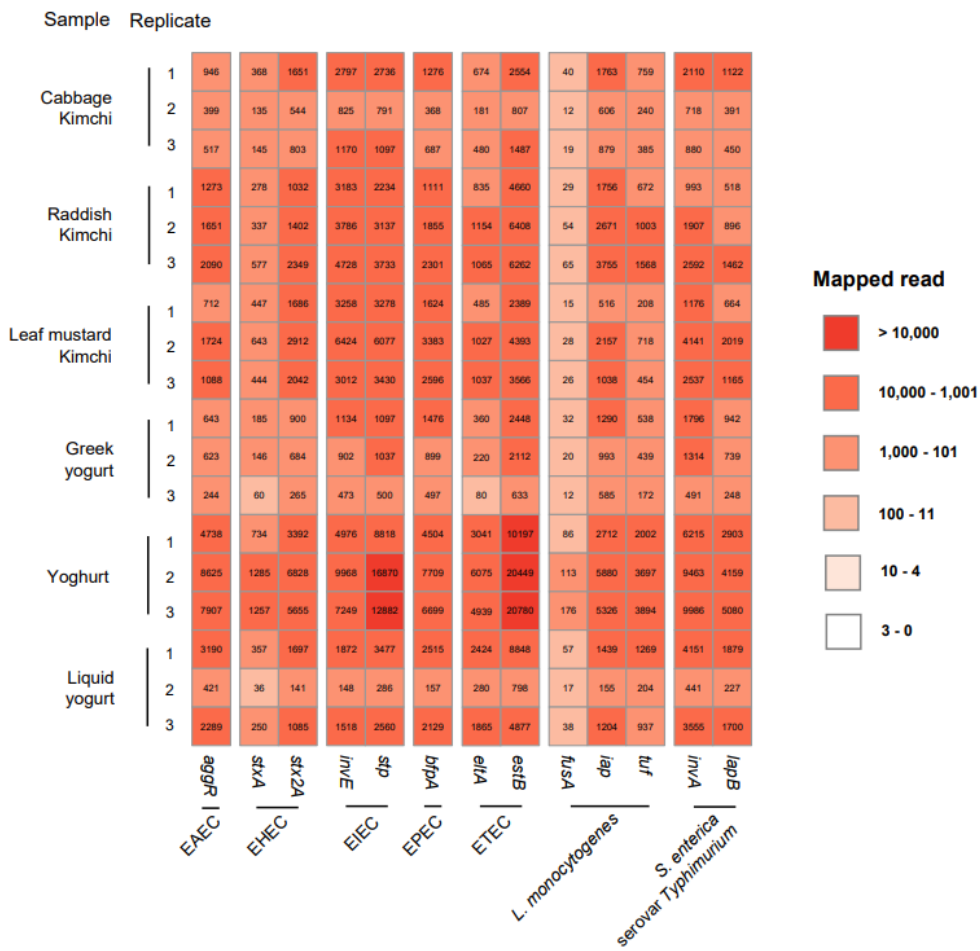

Fig. 21. Cont. Target pathogens-specific genes mapped read in six fermented food samples with or without a mixture of target pathogens.

(C) Six fermented food samples with 10^7 CFU per target pathogen.

D

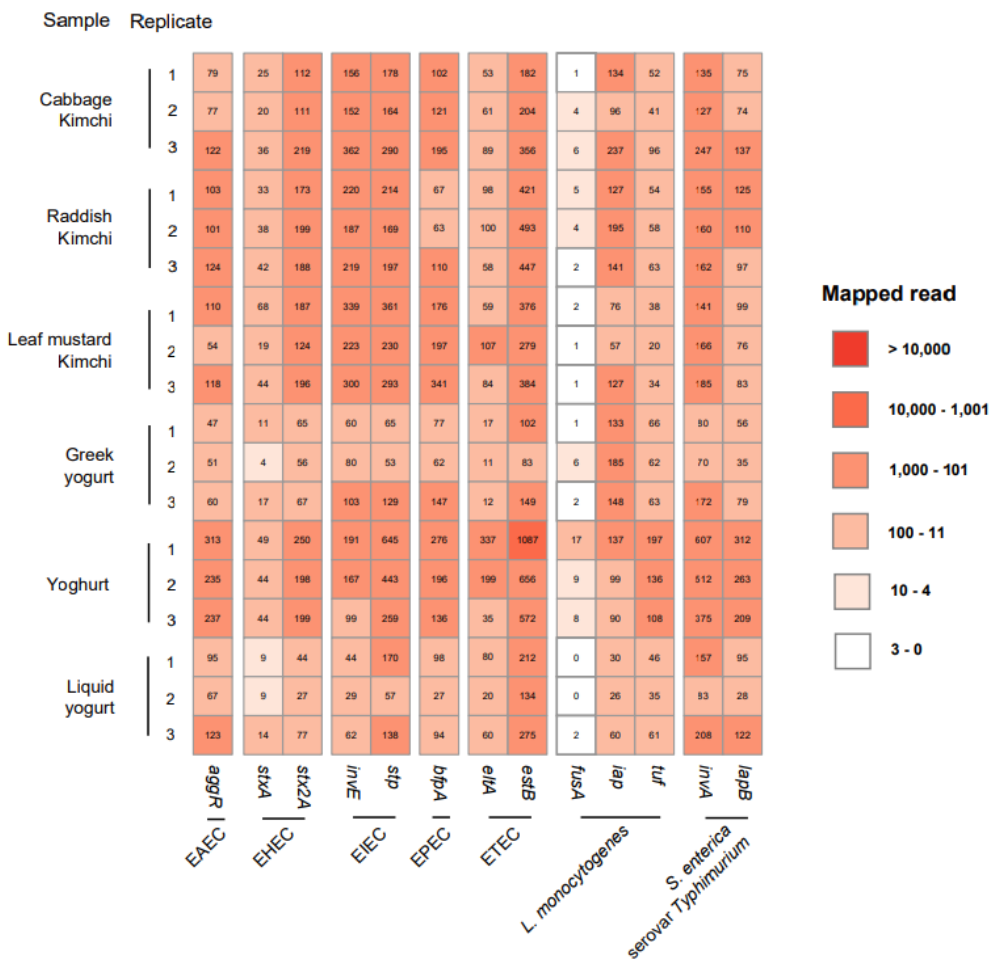

Fig. 21. Cont. Target pathogens-specific genes mapped read in six fermented food samples with or without a mixture of target pathogens.

(D) Six fermented food samples with 10^6 CFU per target pathogen.

E

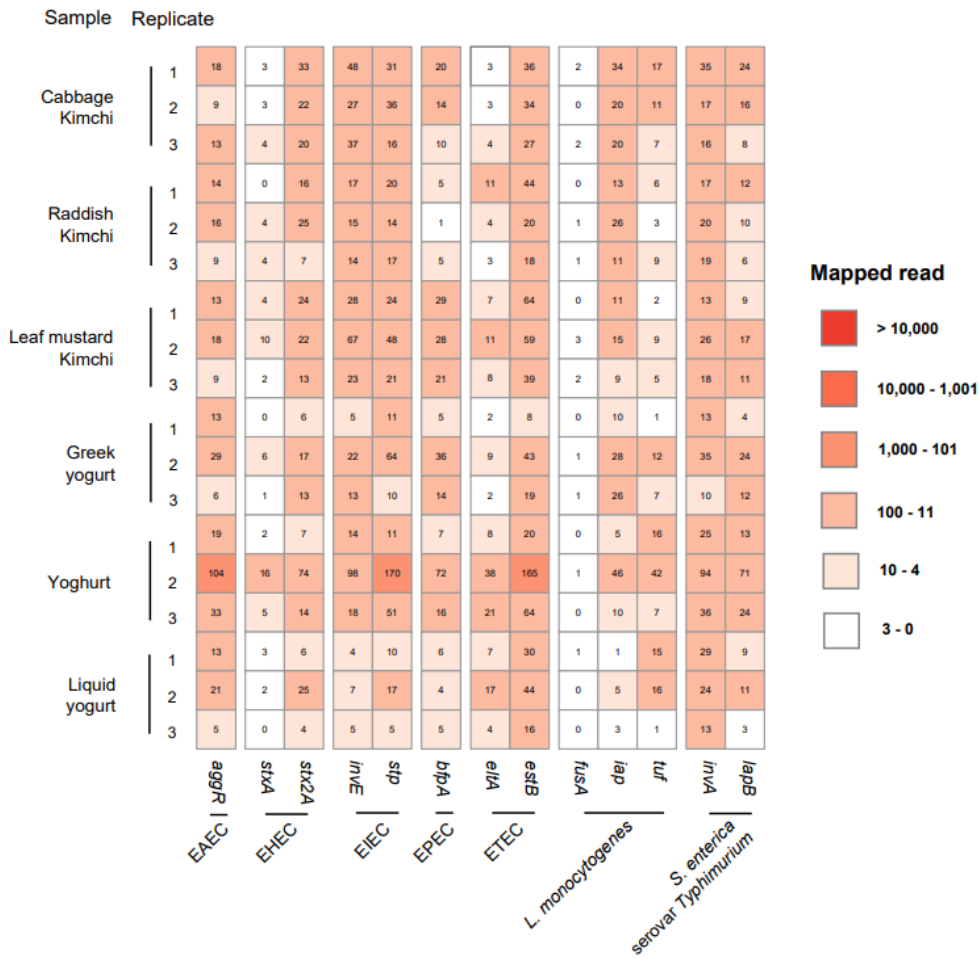

Fig. 21. Cont. Target pathogens-specific genes mapped read in six fermented food samples with or without a mixture of target pathogens.

(E) Six fermented food samples with 10^5 CFU per target pathogen.

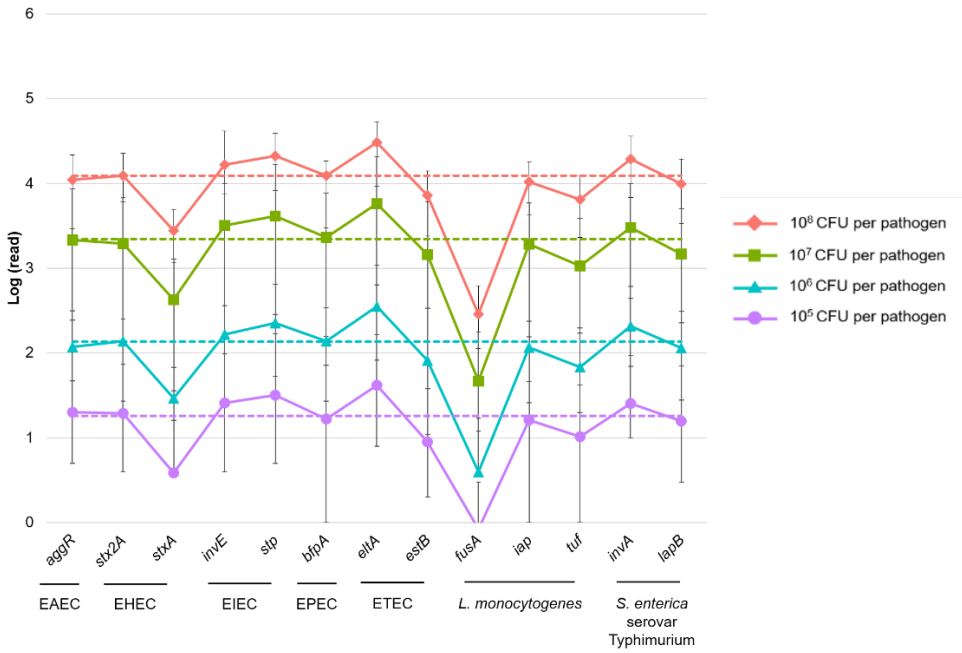

Fig. 22. NGS panel analysis average read summary of target pathogens-specific genes in six fermented food samples with a mixture of target pathogens.

Mapped read counts to target pathogens-specific genes in 10^8, 10^7, 10^6, and 10^5 CFU per target pathogens were indicated in figure as red, green, blue, and purple, respectively, and dashed line indicates total average read in each CFU per target pathogens. Upper error bar means maximum read counts and lower error bar means minimum read counts.

3.2.5. qPCR analysis

qPCR was conducted to compare NGS panel analysis results with qPCR ones for evaluation. Two sets of qPCR DNA templates were prepared and used for qPCR, which were similar to those used for NGS panel analysis. qPCR was performed using three commercial qPCR detection kits, including all detection primer sets targeting seven different species of pathogens. Their qPCR results showed that the average Ct (threshold passed cycle) of target pathogens was 20.89 (10^8 CFU per target pathogen), 24.51 (10^7), 28.17 (10^6), and 31.45 (10^5)(Fig. 23). In addition, Ct and the cell number of target pathogens were negatively proportional, indicating that the rapid detection and identification of target pathogens were associated with low Ct or high cell number of target pathogens (Fig. 23). However, the prepared negative control samples without specific pathogen contamination in samples showed no Ct during the whole qPCR (up to 40 cycles), thereby indicating the absence of target pathogens in negative control samples (Fig. 25A).

Furthermore, Ct values per target pathogen in four dilution factors (10^8, 10^7, 10^6, and 10^5 CFU per target pathogen) were compared to determine the sensitivity and detection limit by qPCR (Fig. 24). In all dilution factors, all target pathogens were fully detected, and they clearly identified six different target pathogens in qPCR without a false-positive (Figs. 25B-E). This result was confirmed in triplicate tests of all fermented food samples.

The highest cell number of target pathogens showed the lowest Ct values, which is consistent with the result shown in Fig. 23. Therefore, the qPCR results indicate that the sensitivity of qPCR may be lower than 10^5 CFU.

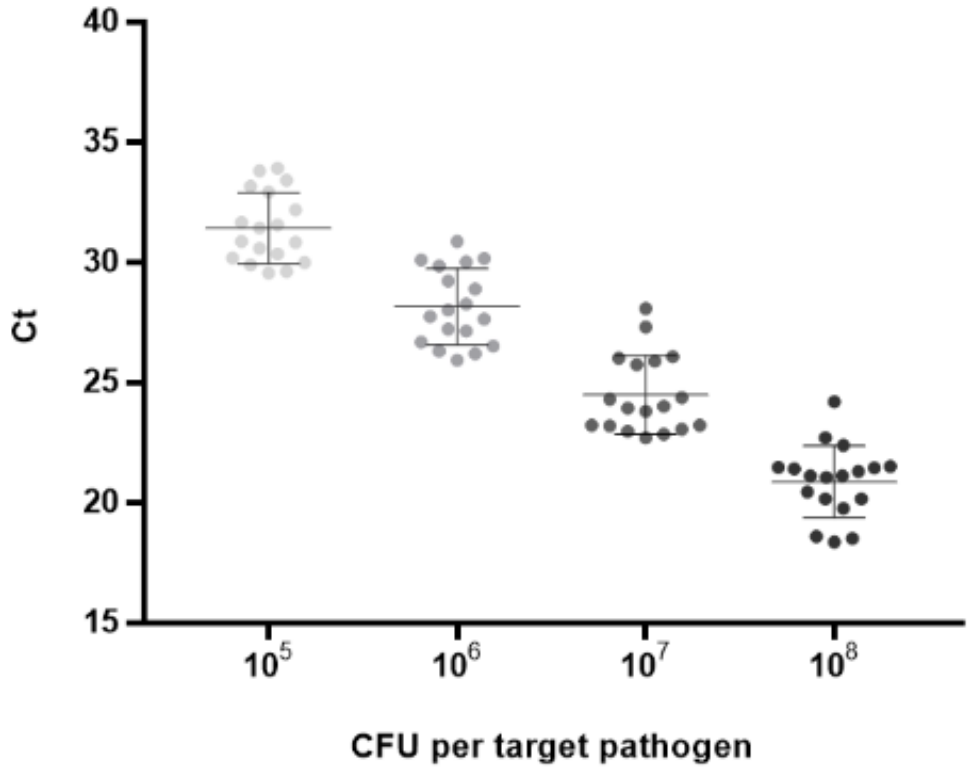

Fig. 23. Dot plot of qPCR results in six fermented food samples with a mixture of target pathogens.

Ct values of 10^8, 10^7, 10^6, and 10^5 CFU per pathogen in six fermented food samples were plotted as dot. Each point represents the mean of target pathogen-specific Ct value.

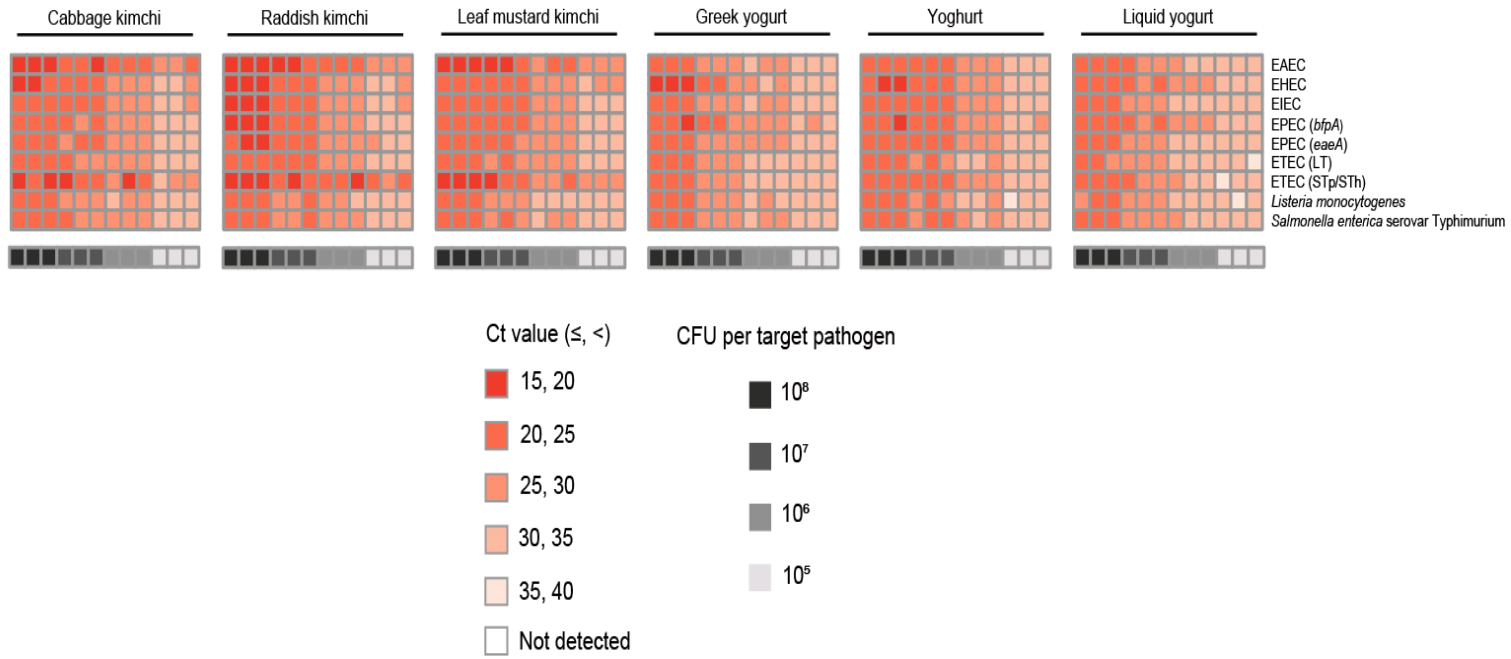

Fig. 24. Heat map of qPCR result in six fermented food samples with a mixture of target pathogens.

Target pathogen-specific gene reads of the qPCR were visualized using heat map. The color-scale of target pathogen-specific gene read or target pathogen Ct value and the level of CFU per target pathogen were indicated below the figure.

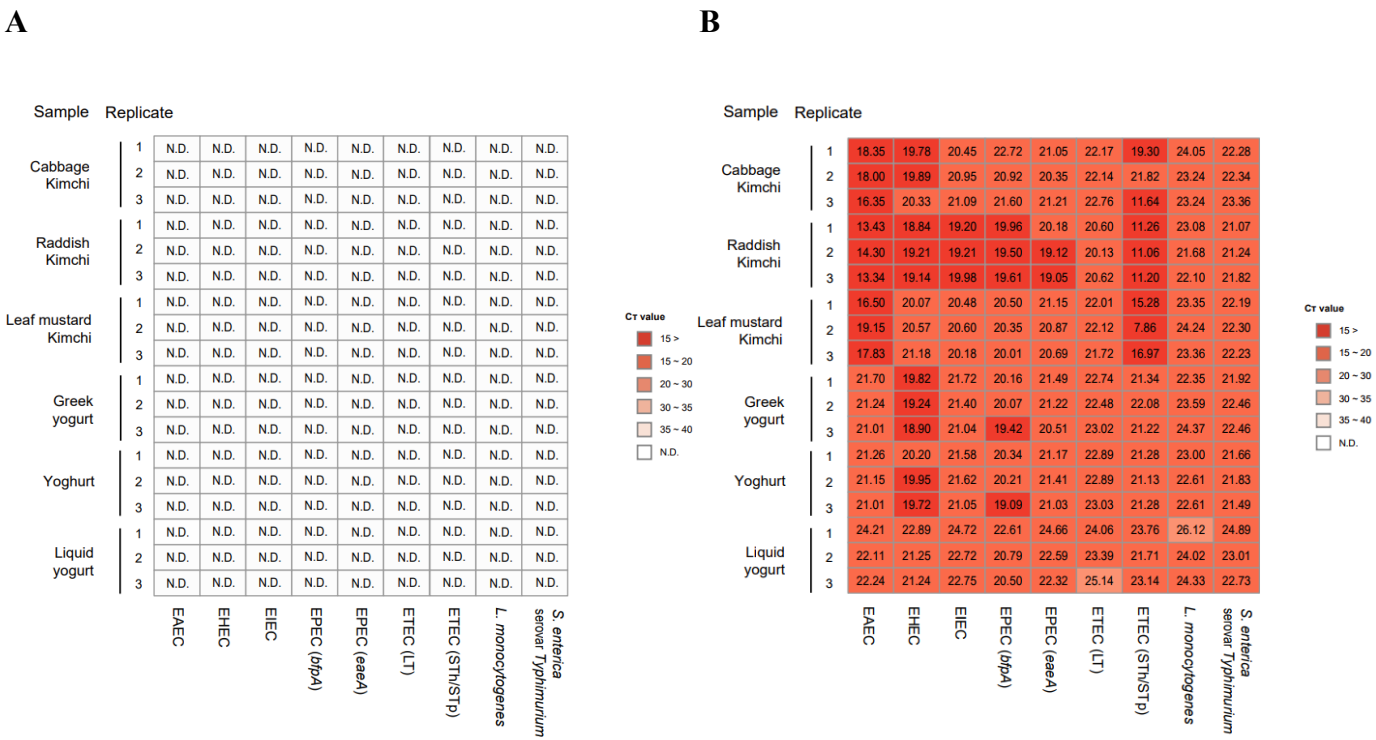

Fig. 25. Ct values of target pathogens in six fermented food samples with or without a mixture of target pathogens. (A) Six fermented food samples without target pathogen, (B) Six fermented food samples with 10^8 CFU per target pathogen. N.D. indicates Ct values were not exceeded threshold until 40 cycles.

C **D**

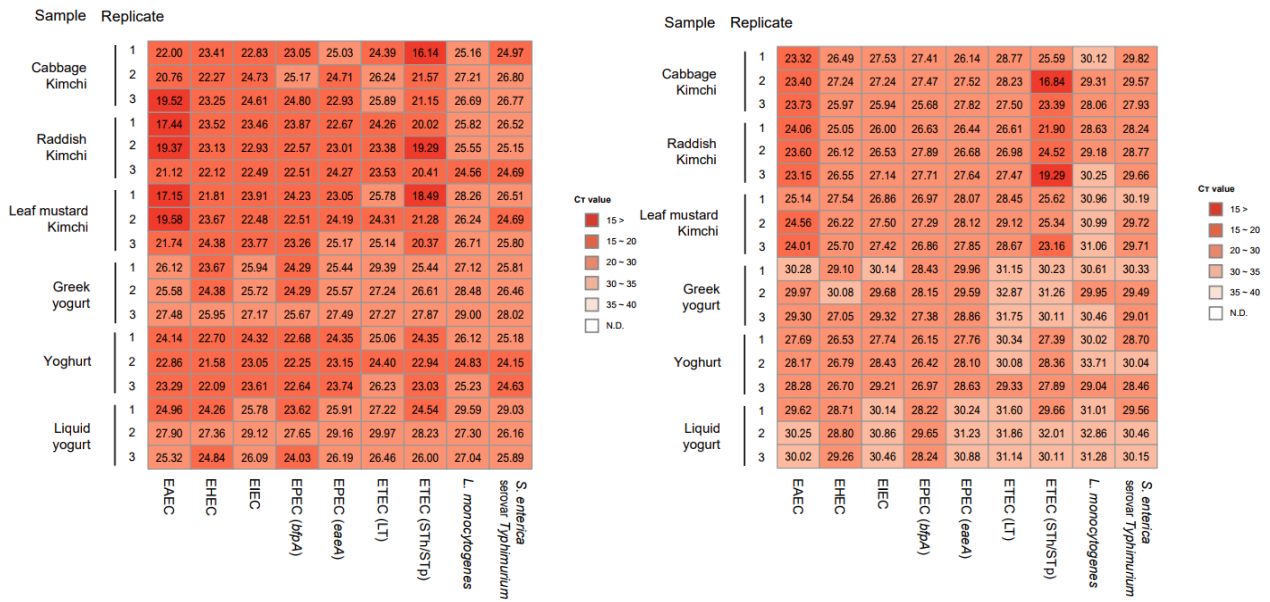

Fig. 25. Cont. Ct values of target pathogens in six fermented food samples with or without a mixture of target pathogens.

(C) Six fermented food samples with 10^7 CFU per target pathogen, (D) Six fermented food samples with 10^6 CFU per target pathogen.

E

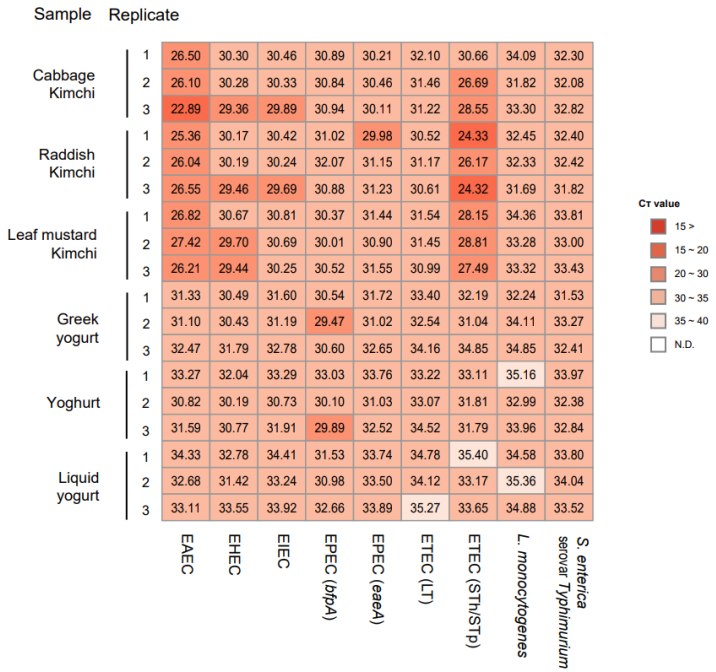

Fig. 25. Cont. Ct values of target pathogens in six fermented food samples with or without a mixture of target pathogens.

(E) Six fermented food samples with 10^5 CFU per target pathogen.

3.2.6. Comparative evaluation of NGS panel analysis and qPCR

NGS panel analysis and qPCR results were compared to evaluate the ability of NGS panel analysis to detect and identify pathogens. Based on previous reports, the qualified read counts and Ct values were correlated with the cell number of target pathogens. Therefore, additional correlation analysis between the qualified read counts and Ct values in each specific target pathogen was performed. In addition, the read counts and Ct values were negatively correlated, and their comparative analysis showed a negative proportional relationship ($y = -12976x + 388696$, $R^2 = 0.5602$; Fig. 26), indicating that the high cell number of target pathogens may be proportional to the rapid detection and identification of pathogens. Moreover, Spearman correlation analysis was performed to statistically compare the results between NGS panel analysis and qPCR in a specific target gene. This analysis revealed negative correlations (Fig. 27). Therefore, this high correlation between the NGS panel and qPCR indicates the importance of the newly developed NGS panel analysis for multiple detection and identification of target pathogens in foods.

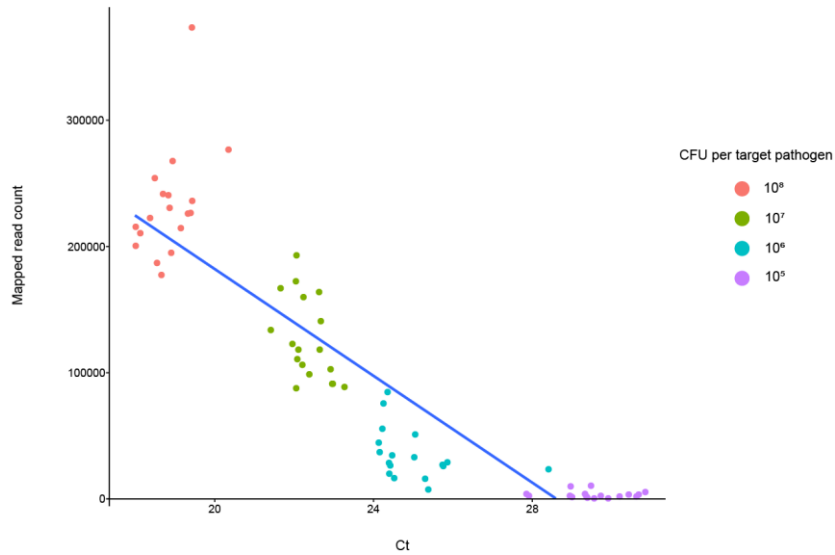

Fig. 26. Linear regression of NGS panel analysis and qPCR results in six fermented food samples with a mixture of target pathogens.

Contaminated samples with 10^8, 10^7, 10^6, and 10^5 CFUs per pathogen were analyzed using NGS panel analysis and qPCR. Each point is the mean of mapped reads to target pathogen-specific gene or Ct values of each target pathogen in a single replicate. Correlation between the average reads mapped to the total target pathogen-specific genes and average total target pathogen Ct values were described in figure.

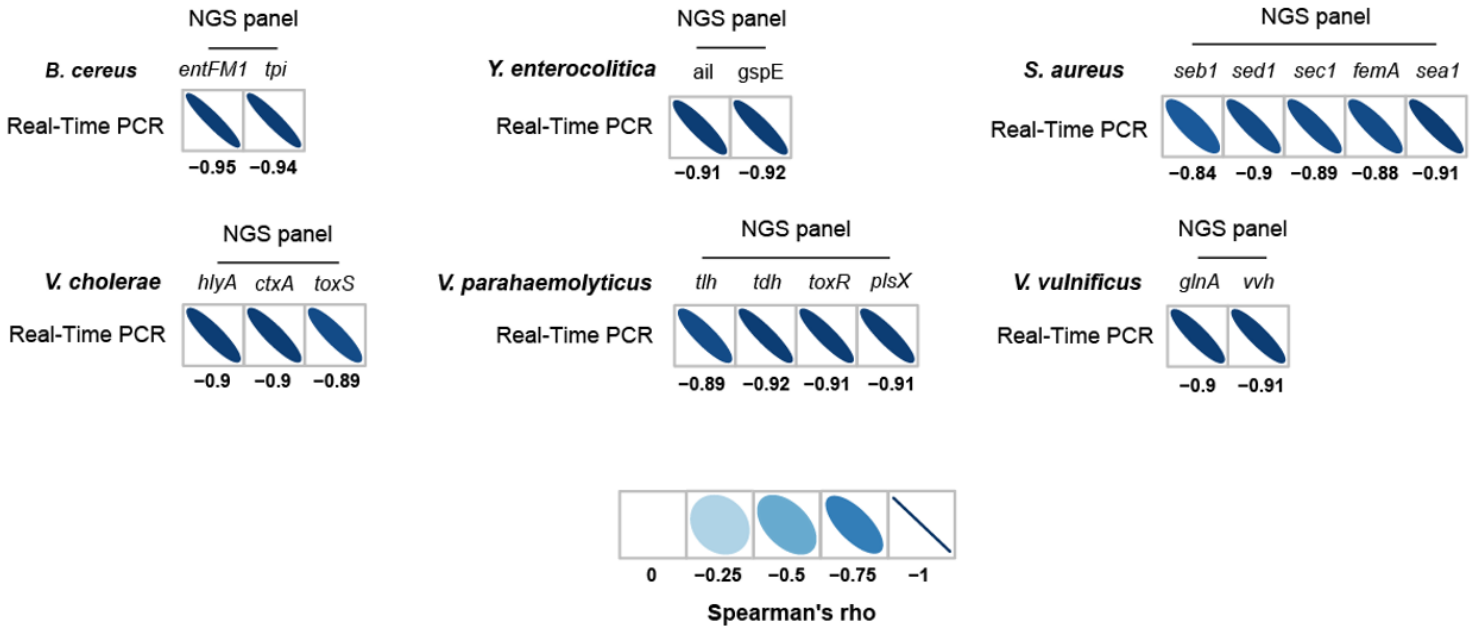

Fig. 27. Spearman correlation between NGS panel analysis and multiplex qPCR results in six fermented food samples with a mixture of target pathogens.

4. Discussion

Foodborne disease outbreaks are generally associated with foodborne pathogen contamination (Newell, Koopmans et al. 2010). However, food environments contain a plethora of food-related microbiota including a variety foodborne pathogen (Lampel, Orlandi et al. 2000); therefore, it is often a challenge to detect and identify the specific foodborne pathogen that has caused an outbreak (Verhoef, Vennema et al. 2010). To detect and identify outbreak-causing pathogens, various methods can be used such as culturing with specific selective media, immune detection using specific antibodies, PCR using foodborne pathogen-targeting specific primer sets, and real-time PCR (Zhao, Lin et al. 2014). However, these methods are not always appropriate for detecting and identifying specific pathogens in foodborne disease outbreaks owing to the complex bacterial ecosystem of food environments (Butz and Patócs 2019). Therefore, new screening methods must be developed and optimized. Accordingly, in the present study, a new method for detecting and identifying multiple pathogens in one reaction, an NGS panel method, was developed, optimized, and compared with a typical qPCR method.

Although the NGS panel method was originally developed for clinical diagnosis and GMO detection (Arulandhu, van Dijk et al. 2018; López-Reig, Fernández-Serra et al. 2019), it has also been used for the multiple detection and identification of foodborne pathogens in food samples (Ferrario, Lugli et al. 2017). In the current study, NGS panel primer sets targeting 18 pathogenic genes were developed and optimized. Using these primer sets, NGS panel analysis was conducted using six agricultural water samples contaminated with pathogens. All pathogens were detected and identified, even with a sample dilution of 10^5 CFUs per pathogen, demonstrating the main advantage of the new method: multiple pathogen detection in one reaction. However, comparative analysis revealed that qPCR has a higher sensitivity than the NGS panel method, although all pathogens could not be detected in one reaction using qPCR. The NGS panel method also gave some false positive results when the number of target pathogen cells was low. Thus, the sensitivity of the NGS panel method when detecting and identifying pathogens must be increased through further optimization of the primer sets. Another major disadvantage of the NGS panel method is the time required to complete NGS sequencing, which could also be optimized through the use of new NGS sequencing technology. For example, nanopore sequencing technology can achieve real-time sequencing (Buytaers, Saltykova et al. 2021) and would therefore be a candidate sequencing method

for minimizing the sequencing time in NGS panel analysis. In summary, although the potential and advantages of the developed NGS panel analysis method were demonstrated in this study, further optimization of the NGS panel primer sets and the application of new real-time NGS sequencing technology will enhance the method's pathogen detection and identification capabilities and help popularize the technology for the improvement of food safety.

5. References

Arulandhu, A. J., van Dijk, J., Staats, M., Hagelaar, R., Voorhuijzen, M., Molenaar, B., et al. (2018). NGS-based amplicon sequencing approach; towards a new era in GMO screening and detection. *Food Control, 93*, 201-210.

Bhunia, A. K. (2014). One day to one hour: how quickly can foodborne pathogens be detected? *Future Microbiology, 9*(8), 935-946.

Bolger, A. M., Lohse, M., Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics, 30*(15), 2114-2120.

Bolton, F. (1998). Strategies in the development of media for the detection of food-borne pathogens. *International Journal of Food Microbiology, 45*(1), 29-34.

Bridier, A. (2019). Exploring foodborne pathogen ecology and antimicrobial resistance in the light of shotgun Metagenomics. *Foodborne Bacterial Pathogens* (pp. 229-245): Springer.

Butz, H., Patócs, A. (2019). Brief summary of the most important molecular genetic methods (PCR, qPCR, microarray, next-generation sequencing, etc.). *Genetics of Endocrine Diseases and Syndromes*, 33-

52.

Buytaers, F. E., Saltykova, A., Denayer, S., Verhaegen, B., Vanneste, K., Roosens, N., et al. (2021). Towards real-time and affordable strain-level metagenomics-based food-borne outbreak investigations using Oxford Nanopore sequencing technologies. *Frontiers in Microbiology*, 3372.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., et al. (2009). BLAST+: architecture and applications. *BMC Bioinformatics, 10*(1), 1-9.

Chen, J., Tang, J., Liu, J., Cai, Z., Bai, X. (2012). Development and evaluation of a multiplex PCR for simultaneous detection of five foodborne pathogens. *Journal of Applied Microbiology, 112*(4), 823-830.

Chung, H. Y., Kim, Y.-T., Kwon, J.-G., Im, H. H., Ko, D., Lee, J.-H., et al. (2021). Molecular interaction between methicillin-resistant Staphylococcus aureus (MRSA) and chicken breast reveals enhancement of pathogenesis and toxicity for food-borne outbreak. *Food Microbiology, 93*, 103602.

De Buyser, M.-L., Dufour, B., Maire, M., Lafarge, V. (2001). Implication of milk and milk products in food-borne diseases in France and in different industrialised countries. *International Journal of Food Microbiology, 67*(1-2), 1-17.

De Magalhães, J. P., Finch, C. E., Janssens, G. (2010). Next-generation sequencing in aging research: emerging applications, problems, pitfalls and possible solutions. *Ageing Research Reviews, 9*(3), 315-323.

Ding, W., Baumdicker, F., Neher, R. A. (2018). panX: pan-genome analysis and exploration. *Nucleic Acids Research, 46*(1), e5-e5.

Elshikh, M. S., Alarjani, K. M., Huessien, D. S., Elnahas, H. A., Esther, A. R. (2022). Enhanced Biodegradation of Chlorpyrifos by Bacillus cereus CP6 and Klebsiella pneumoniae CP19 from municipal waste water. *Environmental Research, 205*, 112438.

Eren, A. M., Esen, Ö. C., Quince, C., Vineis, J. H., Morrison, H. G., Sogin, M. L., et al. (2015). Anvi'o: an advanced analysis and visualization platform for 'omics data. *PeerJ, 3*, e1319.

Ferrario, C., Lugli, G. A., Ossiprandi, M. C., Turroni, F., Milani, C., Duranti, S., et al. (2017). Next generation sequencing-based multigene panel for high throughput detection of food-borne pathogens. *International Journal of Food Microbiology, 256*, 20-29.

Fusco, V., Quero, G. M., Morea, M., Blaiotta, G., Visconti, A. (2011). Rapid and reliable identification of Staphylococcus aureus harbouring the enterotoxin gene cluster (egc) and quantitative detection in raw milk by real time PCR. *International Journal of Food Microbiology, 144*(3),

528-537.

Gupta, A. K., Gupta, U. (2020). Next generation sequencing and its applications. In *Animal Biotechnology* (pp. 395-421): Elsevier.

Hahn, M. A., Keng, P. C., Krauss, T. D. (2008). Flow cytometric analysis to detect pathogens in bacterial cell mixtures using semiconductor quantum dots. *Analytical Chemistry, 80*(3), 864-872.

Joensen, K. G., Scheutz, F., Lund, O., Hasman, H., Kaas, R. S., Nielsen, E. M., et al. (2014). Real-time whole-genome sequencing for routine typing, surveillance, and outbreak detection of verotoxigenic Escherichia coli. *Journal of Clinical Microbiology, 52*(5), 1501-1510.

Lampel, K. A., Orlandi, P. A., Kornegay, L. (2000). Improved template preparation for PCR-based assays for detection of food-borne bacterial pathogens. *Applied and Environmental Microbiology, 66*(10), 4539-4542.

Lee, H., Yoon, Y. (2021). Etiological agents implicated in foodborne illness world wide. *Food Science of Animal Resources, 41*(1), 1.

Lee, J.-H., Ha, J.-H., Lee, H.-W., Lee, J. Y., Hwang, Y.-S., Lee, H. M., et al. (2018). Analysis of microbiological contamination in kimchi and its ingredients. *Journal of Food Hygiene and Safety, 33*(2), 94-101.

Lee, S. H., Yun, J.-W., Lee, J. H., Jung, Y. H., Lee, D. H. (2021). Trends in recent waterborne and foodborne disease outbreaks in South Korea,

2015–2019. *Osong Public Health and Research Perspectives, 12*(2), 73.

Lee, W.-C., Lee, M.-J., Kim, J.-S., Park, S.-Y. (2001). Foodborne illness outbreaks in Korea and Japan studied retrospectively. *Journal of Food Protection, 64*(6), 899-902.

Liu, B., Zheng, D., Zhou, S., Chen, L., Yang, J. (2022). VFDB 2022: a general classification scheme for bacterial virulence factors. *Nucleic Acids Research, 50*(D1), D912-D917.

Liu, X., Liu, G., Wu, Y., Pang, X., Wu, Y., Niu, J., et al. (2021). Transposon sequencing: A powerful tool for the functional genomic study of food-borne pathogens. *Trends in Food Science & Technology, 118*, 679-687.

Liu, Y., Cao, Y., Wang, T., Dong, Q., Li, J., Niu, C. (2019). Detection of 12 common food-borne bacterial pathogens by TaqMan real-time PCR using a single set of reaction conditions. *Frontiers in Microbiology, 10*, 222.

López-Reig, R., Fernández-Serra, A., Romero, I., Zorrero, C., Illueca, C., García-Casado, Z., et al. (2019). Prognostic classification of endometrial cancer using a molecular approach based on a twelve-gene NGS panel. *Scientific Reports, 9*(1), 1-9.

Lorenz, T. C. (2012). Polymerase chain reaction: basic protocol plus troubleshooting and optimization strategies. *JoVE (Journal of*

Visualized Experiments)(63), e3998.

Masella, A. P., Bartram, A. K., Truszkowski, J. M., Brown, D. G., Neufeld, J. D. (2012). PANDAseq: paired-end assembler for illumina sequences. *BMC Bioinformatics, 13*(1), 1-7.

Mira Miralles, M., Maestre-Carballa, L., Lluesma-Gomez, M., Martinez-Garcia, M. (2019). High-throughput 16S rRNA sequencing to assess potentially active bacteria and foodborne pathogens: a case example in ready-to-eat food. *Foods, 8*(10), 480.

Müller, D., Greune, L., Heusipp, G., Karch, H., Fruth, A., Tschäpe, H., et al. (2007). Identification of unconventional intestinal pathogenic Escherichia coli isolates expressing intermediate virulence factor profiles by using a novel single-step multiplex PCR. *Applied and Environmental Microbiology, 73*(10), 3380-3390.

Newell, D. G., Koopmans, M., Verhoef, L., Duizer, E., Aidara-Kane, A., Sprong, H., et al. (2010). Food-borne diseases—the challenges of 20 years ago still persist while new ones continue to emerge. *International Journal of Food Microbiology, 139*, S3-S15.

Oh, S.-Y., Nam, K.-W., Yoon, D.-H. (2017). Analysis of pathogenic microorganism's contamination and heavy metals on kimchi cabbage by cultivation methods in Korea. *Journal of Food Hygiene and Safety, 32*(6), 500-506.

Palka-Santini, M., Cleven, B. E., Eichinger, L., Krönke, M., Krut, O. (2009). Large scale multiplex PCR improves pathogen detection by DNA microarrays. *BMC Microbiology, 9*(1), 1-14.

Park, M., Won, J., Choi, B. Y., Lee, C. J. (2020). Optimization of primer sets and detection protocols for SARS-CoV-2 of coronavirus disease 2019 (COVID-19) using PCR and real-time PCR. *Experimental & Molecular Medicine, 52*(6), 963-977.

Pianetti, A., Sabatini, L., Bruscolini, F., Chiaverini, F., Cecchetti, G. (2004). Faecal contamination indicators, Salmonella, Vibrio and Aeromonas in water used for the irrigation of agricultural products. *Epidemiology & Infection, 132*(2), 231-238.

Priyanka, B., Patil, R. K., Dwarakanath, S. (2016). A review on detection methods used for foodborne pathogens. *The Indian Journal of Medical Research, 144*(3), 327.

Roulová, N., Moťková, P., Brožková, I., Swiontek Brzezinska, M., Pejchalová, M. (2022). Detection, characterization, and antimicrobial susceptibility of Yersinia enterocolitica in different types of wastewater in the Czech Republic. *Journal of Applied Microbiology*.

Shi, L., Wu, F., Wen, Y., Zhao, F., Xiang, J., Ma, L. (2015). A novel method to detect Listeria monocytogenes via superparamagnetic lateral flow immunoassay. *Analytical and Bioanalytical Chemistry, 407*(2), 529-

535.

Shim, W.-B., Choi, J.-G., Kim, J.-Y., Yang, Z.-Y., Lee, K.-H., Kim, M.-G., et al. (2008). Enhanced rapidity for qualitative detection of Listeria monocytogenes using an enzyme-linked immunosorbent assay and immunochromatography strip test combined with immunomagnetic bead separation. *Journal of Food Protection, 71*(4), 781-789.

Silva, V., Caniça, M., Capelo, J. L., Igrejas, G., Poeta, P. (2020). Diversity and genetic lineages of environmental staphylococci: A surface water overview. *FEMS Microbiology Ecology, 96*(12), fiaa191.

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www. R-project. org/. (Accessed August 2022)

Tokarskyy, O., Marshall, D. L. (2008). Immunosensors for rapid detection of Escherichia coli O157: H7—Perspectives for use in the meat processing industry. *Food Microbiology, 25*(1), 1-12.

Umesha, S., Manukumar, H. (2018). Advanced molecular diagnostic techniques for detection of food-borne pathogens: Current applications and future challenges. *Critical Reviews in Food Science and Nutrition, 58*(1), 84-104.

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M., et al. (2012). Primer3—new capabilities and interfaces. *Nucleic*

Acids Research, 40(15), e115-e115.

Vega, M., Lerner, B., Lasorsa, C. A., Pierpauli, K., Perez, M. S. (2016). Automated and low cost method to manufacture addressable solid-state nanopores. *Microsystem Technologies, 22*(1), 109-117.

Verhoef, L., Vennema, H., Van Pelt, W., Lees, D., Boshuizen, H., Henshilwood, K., et al. (2010). Use of norovirus genotype profiles to differentiate origins of foodborne outbreaks. *Emerging Infectious Diseases, 16*(4), 617.

Yang, Y.-W., Chen, M.-K., Yang, B.-Y., Huang, X.-J., Zhang, X.-R., He, L.-Q., et al. (2015). Use of 16S rRNA gene-targeted group-specific primers for real-time PCR analysis of predominant bacteria in mouse feces. *Applied and Environmental Microbiology, 81*(19), 6749-6756.

Yohe, S., Thyagarajan, B. (2017). Review of clinical next-generation sequencing. *Archives of Pathology & Laboratory Medicine, 141*(11), 1544-1557.

Zakotnik, S., Knap, N., Bogovič, P., Zorec, T. M., Poljak, M., Strle, F., et al. (2022). Complete Genome Sequencing of Tick-Borne Encephalitis Virus Directly from Clinical Samples: Comparison of Shotgun Metagenomic and Targeted Amplicon-Based Sequencing. *Viruses, 14*(6), 1267.

Zhao, X., Lin, C.-W., Wang, J., Oh, D. H. (2014). Advances in rapid detection

methods for foodborne pathogens. *Journal of Microbiology and Biotechnology, 24*(3), 297-312.

국문초록

복잡한 균 총을 가지고 있는 식품 속에서 식중독 사고의 원인 균 만을 감지하고 식별하는 것은 중요하다. 현재까지 사용되고 있는 식중독균 검출 및 식별 기술은 위와 같은 목표를 달성하기 위해 여러 문제점들을 해결해왔지만, 동시다발적으로 다양한 식중독균을 검출하는데 있어서 한계점을 나타냈다. 따라서, 한 번의 반응으로 다양한 식중독 원인 균을 효율적으로 선별하고 식별할 수 있다고 보고된 NGS 패널 기술을 활용하여 새로운 식중독균 검출 및 식별 기술을 개발하였다.

본 연구에서는 2 가지의 NGS 패널로 각각 6 종 및 7 종의 식중독균 (세트1: *Bacillus cereus*, *Yersinia enterocolitica*, *Staphylococcus aureus*, *Vibrio cholerae*, *Vibrio parahaemolyticus* 및 *Vibrio vulnificus*, 세트2: *Listeria monocytogenes*, *Salmonella enterica* serovar Typhimurium, enteropathogenic *Escherichia coli* [EPEC], enteroinvasive *E. coli* [EIEC], enterotoxigenic *E. coli* [ETEC], enterohemorrhagic *E. coli* [EHEC], 및 enteroaggregative *E.*

coli [EAEC clinical (EAEC)])안에서 18 개 및 13 개의 특이적인 독성 인자 유전자를 표적으로 하는 새로운 NGS 패널 프라이머 세트를 개발하고 최적화했다. 프라이머 세트를 이용한 싱글플렉스 PCR에서는 예측된 크기의 단일 PCR 앰플리콘이 나타났고, 이후의 교차 확인 및 멀티플렉스 PCR에서는 비특이적 프라이머 세트 또는 비특이적 식중독 균의 DNA에 의한 간섭이 나타나지 않아 새로운 프라이머 세트의 특이성과 선택성을 확인했다.

이후, 새로운 NGS 패널 방법의 평가를 위해 수집된 6개의 서로 다른 농업용수 샘플과 6개의 서로 다른 발효식품에 각각 6 종과 7 종의 식중독균을 동시 오염시킨 후 NGS 패널 분석을 진행하였다. 그 결과, 농업용수에서는 10^8~10^5 CFUs 수준에서 18 개의 표적 유전자가, 발효식품에서는 식중독균 종 당 10^8~10^7 CFUs 수준에서 13 개의 표적 유전자가 다중 검출 및 식별되었다. 또한, 독성 인자 유전자의 평균 총 서열 판독 횟수는 표적 병원체 당 CFU와 양의 상관관계가 있었다. 하지만, NGS 패널 분석은 한 반응에서 다양한 종의 식중독균을 동시 검출하는 이점을 보여주었지만, 적은 CFU (희석 계수 10^6−10^5)의 식중독균이 오염된 샘플에서 상대적으로 낮은 감도와 위양성 결과가 발생했다. 추가적으

로, NGS 패널 결과를 검증하기 위해 동일한 오염된 농업용수 및 발효식품 샘플을 사용하여 두가지 세트 및 세가지 세트의 qPCR 분석을 수행했으며, 표적 병원체 검출 및 식별의 효율성 및 특이성은 NGS 패널 분석과 유사했다. 비교 통계 분석 및 Spearman 상관 분석은 NGS 패널 서열 판독 횟수와 qPCR 주기 임계값(Ct) 값이 음의 연관이 있음을 보여주었으며 결과의 유사성을 나타내었다.

보다 빠르고 정확한 검출 및 식별을 위해 NGS 패널 분석을 향상시키려면 NGS 패널 프라이머 세트의 추가적인 최적화와 실시간 NGS 시퀀싱 기술의 도입이 필요하다. 결과적으로, 위 연구를 통해 식품 매개 병원체의 다중 검출을 위한 NGS 패널 분석의 적용에 대한 잠재력과 이점들을 확인하였다.