A THESIS FOR THE DEGREE OF MASTER OF SCIENCE

Development of a Novel O-serotyping Method for the Rapid Detection of Various *Escherichia coli* O-serotypes Using Multiplex Real-Time PCR

다중 실시간 중합효소연쇄반응을 이용해
대장균 O-혈청형의 신속한 검출을 위한
새로운 동정 방법 개발

February, 2023

Eun-Ji Shin

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

Development of a Novel O-serotyping Method for the Rapid Detection of Various *Escherichia coli* O-serotypes Using Multiplex Real-Time PCR

지도교수 이 주 훈

이 논문을 농학석사 학위논문으로 제출함

2023년 2월

서울대학교 대학원

농생명공학부

신 은 지

신은지의 석사 학위논문을 인준함

2023년 2월

위 원 장 유 상 열 (인)

부 위 원 장 이 주 훈 (인)

위 원 최 상 호 (인)

ABSTRACT

Shin, Eun-Ji

Department of Agricultural Biotechnology

The Graduate School

Seoul National University

Foodborne outbreaks caused by pathogenic *E. coli* are worldwide occurrence that cause illness, economic losses, and even fatalities. Previous studies have shown that *E. coli* pathogenicity and serotypes are highly correlated; thus, the rapid and accurate identification of *E. coli* serotypes is essential for preventing severe foodborne outbreaks. In this study, a novel multiplex real-time PCR method was developed for *E. coli* O-serotyping that overcame some limitations of conventional and *in silico* serotyping, i.e., it performed with high sensitivity, high specificity, and a low detection limit and reduced O-serotype detection and identification time without any loss of sensitivity and specificity. Specifically, two O-serotype-specific genes, *wzy* and *wzx*, were used to design 21 primer/probe sets for determining 21 different O-serotypes. The method was then evaluated using 11 O-serotype reference strains and 43 environmental isolates: O-serotyping was completed in

<40 min with a low detection limit of 1–10 pg DNA. In addition, the PCR serotyping method showed the same results as the existing serotyping method, confirming that the method was effective. Importantly, a food application test was also completed: the method could identify O-serotypes in ground beef samples even at 10^1 CFU per sample. Based on our results, our novel method can be considered the optimal O-serotyping tool for the determination of *E. coli* O-serotypes produced to date. The developed multiplex real-time PCR serotyping method can rapidly, inexpensively, and accurately identify *E. coli* O-serotypes; therefore, it could help prevent foodborne outbreaks.

***Keywords*: Real-time PCR, O-serotype, *Escherichia coli*, Serotyping, Multiplex PCR**

Student Number: 2021-29876

CONTENTS

List of Figures

List of Tables

1. Introduction

Escherichia coli has been found in various environmental habitats, including human and animal intestines, water, soil, and even some foods. Some of these *E. coli* strains have been reported as the principal causative agents of foodborne diseases (Kaper, Nataro, & Mobley, 2004). Foodborne outbreaks occur when food contaminated with pathogenic *E. coli* is ingested. Although *E. coli* is generally nonpathogenic, certain serotypes have been associated with foodborne outbreaks of diseases with symptoms such as mild diarrhea, severe hemorrhagic colitis, and hemolytic uremic syndrome (HUS) (Chen & Hsueh, 2011). One representative case of an *E. coli* foodborne outbreak involved *E. coli* serotype O104:H4 and caused 53 confirmed deaths in Europe (51 of which occurred in Germany) in summer 2011 (Burger, 2012). In the United States, multistate outbreaks were caused by romaine lettuce contaminated with *E. coli* serotype O157:H7 in 2018; 96 of 210 infected people were hospitalized, including 27 people with HUS, and 5 deaths were reported (Baber ISWJ, 2018). In Victoria in Canada, a foodborne outbreak was caused by sandwiches containing raw minced celery contaminated with Shiga toxin-producing *E. coli* (STEC) O103 in

April 2021 (Smith, Griffiths, Allison, Hoyano, & Hoang, 2022); this was the first non-O157 outbreak in Canada, causing five hospitalizations and one death in six cases. According to the National Outbreak Reporting System of the US Centers for Disease Control and Prevention, eight major *E. coli* O-serotypes (O157, O26, O111, O121, O103, O145, O45, and O6) were involved in clinical outbreaks that occurred in 2008–2017 (Dewey-Mattia, Manikonda, Hall, Wise, & Crowe, 2018; Lee et al., 2005), indicating that specific *E. coli* O-serotypes are associated with foodborne outbreaks; thus, the rapid identification of such O-serotypes could help prevent severe foodborne outbreaks, such as those caused by O104:H4. To this end, a rapid *E. coli* O-serotype identification method with high accuracy and sensitivity is required.

The O-antigen forms a part of the lipopolysaccharide (LPS) in the outer membrane of gram-negative bacteria. The O-antigen is a target point for bacteriophage infection as well as the innate and adaptive immune system; therefore, it is thought to be involved in host–pathogen interactions, particularly resistance to serum-mediated killing and phagocytosis (Bazaka, Crawford, Nazarenko, & Ivanova, 2011; Reeves, 1995). In addition, the structure and length of the O-antigen has been associated with determining the pathogenicity of the host bacterium

(West et al., 2005). Therefore, the specific O-antigen in the outer membrane of *E. coli* may be an important virulence factor for its pathogenicity (Cho, Oh, Kim, Oh, & Park, 2010; Tamura, Sakazaki, Murase, & Kosako, 1996), supporting the importance of identifying *E. coli* O-serotypes. Based on the polymorphic trait of the O-antigen of *E. coli,* 186 O-serotypes have been reported (Fratamico, DebRoy, Liu, Needleman, Baranzoni, & Feng, 2016; Sun et al., 2011), further highlighting the need for a rapid and accurate *E. coli* O-serotype identification method to ascertain the source of an outbreak and monitor and prevent food poisoning caused by pathogenic *E. coli* with specific O-antigens.

O-antigen identification via serotyping is important for classifying subgroups within a single microbial species according to the distinctive cell surface antigens associated with the immune response. Kauffmann (1947) suggested a taxonomy of *E. coli* based on antigen characteristics, and *E. coli* serotype identification was subsequently developed based on a combination of three immunogenic components: somatic cell O-antigens, flagellar H-antigens, and capsular K-antigens (Ørskov & Ørskov, 1984). Conventional O-serotyping based on the agglutination reaction between somatic cell O-antigens and antiserum

produced by immunizing rabbits with boiled cultures has been considered the standard method of serotype identification. However, this method is time-consuming, labor-intensive, and expensive, and serotyping results are not clearly determined in some cases (X. Wang & Quinn, 2010). Given the increased availability of *E. coli* genome sequences in public databases, *in silico* serotyping methods have been developed to determine *E. coli* serotypes. Such serotyping is based on comparing the O-antigen-related gene sequences of *E. coli* whole genomes. Although O-serotype identification is rapid and highly accurate and an *in silico* serotyping program, SerotypeFinder 2.0 (Joensen, Tetzschner, Iguchi, Aarestrup, & Scheutz, 2015), is freely available, this serotyping method is dependent on the availability of the whole genome sequence of the target *E. coli* strain. However, obtaining this genome sequence via next-generation sequencing (NGS) may be expensive and time-consuming (Joensen et al., 2015). Therefore, an inexpensive and rapid serotyping method with high accuracy and fidelity must be developed to overcome these disadvantages.

The results of previous studies have suggested that some *E. coli* serotypes can be identified using an O-antigen gene–based PCR method. Indeed, a PCR-based serotyping method was developed using *wzx–wzy*

gene sequences and evaluated though the identification of non-O157 Shiga-toxigenic *E. coli* (STEC) O26 (DebRoy, Fratamico, Roberts, Davis, & Liu, 2005). Subsequently, a PCR serotyping method using O-antigen biosynthetic genes (*wzy*, *wzx*, *wzt*, and *wzm*) was developed to identify 147 individual *E. coli* O-serotypes (Iguchi, Iyoda, Seto, Morita-Ishihara, Scheutz, & Ohnishi, 2015), based on which a multiplex PCR serotyping method with 20 multisets contain 6–9 primer pairs was also developed to identify O-serotypes with different sizes of PCR amplicons. These studies confirmed that PCR serotyping is an accurate and inexpensive serotyping method. However, despite its advantages, performing the PCR serotyping method requires additional hours for agarose gel electrophoresis analysis and manual determination, i.e., it is not a real-time detection method. To overcome this disadvantage, real-time PCR could be used to improve the speed of *E. coli* serotype detection and identification; however, real-time PCR–based serotyping and related optimized PCR primer sets have not been developed.

In the present study, to meet the need for real-time PCR–based serotyping, *E. coli* serotyping PCR primer sets for real-time PCR were developed using O-antigen genes (*wzy* and *wzx*), and their specificity and accuracy were evaluated using singleplex, cross-check, and multiplex

PCRs with *E. coli* strains with different O-serotypes. Having developed and evaluated the serotyping PCR primer sets, a real-time PCR method employing the serotyping primer sets and associated probes was developed and optimized for real-time PCR serotyping, and this method was evaluated using *E. coli* strains with different serotypes. Our results indicate that this new real-time PCR serotyping method can rapidly, accurately, and inexpensively identify the serotypes of *E. coli* to improve food safety.

2. Materials and Methods

2.1. Bacterial strains, culture media, and growth conditions

Eleven *E. coli* reference strains with O-serotype information and 43 *E. coli* isolates without O-serotype information were used in this study and are listed in Table 1 and 2, respectively. To obtain these *E. coli* isolates, 5 g of each collected sample was mixed with 45 ml of 0.1% sterilized peptone water and transferred into 3M sterile sample bags (USA). The mixture was stomached using a BagMixer CC (Interscience, France). The supernatant was 10-fold serially diluted with 0.1% sterilized peptone water and then spread on 5-bromo-4-chloro-3-indolyl-β-d-glucuronide agar medium (BCIG; MBcell, Korea) as a selective medium. After incubation at 37°C for 12 h, colonies were randomly picked and cultured in fresh Luria–Bertani (LB; Difco, USA) broth medium with shaking at 37°C.

Table 1. *E. coli* reference strains and their O-serotypes

Strain	Source	Serotype	Reference[a]
NCCP 15954	Stool	O145	NCCP
ATCC 43895	Raw hamburger meat	O157	ATCC
NCCP 15961	Stool	O26	NCCP
NCCP 15956	Stool	O103	NCCP
ATCC 43887	Feces	O111	ATCC
NCCP 15732	Stool	O6	NCCP
NCCP 16186	Stool	O25	NCCP
ATCC 35401	Feces	O78	ATCC
NCCP 13987	Stool	O55	NCCP
NCCP 15654	Stool	O104	NCCP
ATCC 43893	Feces	O124	ATCC

[a], NCCP, National Culture Collection for Pathogens, ATCC, American Type Culture Collection

Table 2. *E. coli* isolates and their genome sequence information

	Strain	Source	Accession number
1	SNU-Eco001	Environment	JAOWSX000000000
2	SNU-Eco002	Environment	JAOWSW000000000
3	SNU-Eco003	Environment	JAOWSV000000000
4	SNU-Eco004	Environment	JAOWSU000000000
5	SNU-Eco005	Environment	JAOWST000000000
6	SNU-Eco006	Environment	JAOWSS000000000
7	SNU-Eco007	Environment	JAOWSR000000000
8	SNU-Eco008	Environment	JAOWSQ000000000
9	SNU-Eco009	Environment	JAOWSP000000000
10	SNU-Eco010	Environment	JAOWSO000000000
11	SNU-Eco011	Clinical	JAOWSN000000000
12	SNU-Eco012	Clinical	JAOWSM000000000
13	SNU-Eco013	Chicken	JAOWSL000000000
14	SNU-Eco014	Environment	JAOWSK000000000
15	SNU-Eco015	Environment	JAOWSJ000000000
16	SNU-Eco016	Environment	JAOWSI000000000
17	SNU-Eco017	Environment	JAOWSH000000000
18	SNU-Eco018	Environment	JAOWSG000000000
19	SNU-Eco019	Environment	JAOWSF000000000
20	SNU-Eco020	Environment	JAOWSE000000000
21	SNU-Eco021	Environment	JAOWSD000000000
22	SNU-Eco022	Environment	JAOWSC000000000
23	SNU-Eco023	Environment	JAOWSB000000000
24	SNU-Eco024	Environment	JAOWSA000000000
25	SNU-Eco025	Environment	JAOWRZ000000000
26	SNU-Eco026	Environment	JAOWRY000000000
27	SNU-Eco027	Environment	JAOWRX000000000
28	SNU-Eco028	Environment	JAOWRW000000000
29	SNU-Eco029	Environment	JAOWRV000000000
30	SNU-Eco030	Environment	JAOWRU000000000
31	SNU-Eco031	Environment	JAOWRT000000000
32	SNU-Eco032	Environment	JAOWRS000000000
33	SNU-Eco033	Environment	JAOWRR000000000
34	SNU-Eco034	Environment	JAOWRQ000000000
35	SNU-Eco035	Environment	JAOWRP000000000
36	SNU-Eco036	Pig fecal	JAOWRO000000000
37	SNU-Eco037	Cow fecal	JAOWRN000000000
38	SNU-Eco038	Spinach	JAOWRM000000000
39	SNU-Eco039	Cow fecal	JAOWRL000000000
40	SNU-Eco040	Cow fecal	JAOWRK000000000
41	SNU-Eco041	Cow fecal	JAOWRJ000000000
42	SNU-Eco042	Cow fecal	JAOWRI000000000
43	SNU-Eco043	Cow fecal	JAOWRH000000000

2.2. Genomic DNA extraction

E. coli genomic DNA was extracted and purified using a G-spin™ Genomic DNA Extraction Kit for Bacteria (INtRON, Korea) following the manufacturer's standard protocols. A NanoDrop™ 2000 Spectrophotometer (Thermo Fisher Scientific, USA) was used to determine the purity and quantity of extracted genomic DNA.

2.3. 16S rRNA gene sequencing

The 16S rRNA gene was amplified using PCR with a 16S rRNA universal primer set: 27F (5′-AGAGTTTGATCCTGGCTCAG-3′) and 1492R (5′-GGTTACCTTGTTACGACTT-3′) (Lane, 1991). The PCR reaction mixture consisted of 1 µl of template DNA (40 ng/µl), 0.75 µl of forward and reverse primers (20 µM each), 2.5 µl of 10X *Taq* buffer, 2.5 µl of dNTP mixture (2 mM each; MGMed, Korea), and 0.5 µl of *Taq* DNA polymerase (1 U/µl; MGMed), and the total reaction volume was adjusted to 25 µl using Ultrapure Water (Welgene, Korea). The PCR conditions were as follows: 1 cycle of 94°C for 3 min; 35 cycles of 94°C for 30 s, 62°C for 1 min, and 72°C for 1 min; and 1 cycle of 72°C for 5 min. After PCR amplification, the PCR amplicon was purified using a NICSROprep™ PCR Clean-up S & V Kit (Bionics, Korea) and sequenced using capillary sequencing (Bionics). The *E. coli* isolates

were identified using 16S rRNA gene sequence analysis via NCBI BLASTN (USA).

2.4. NGS-based genome sequencing

The genome library was prepared using a TruSeq DNA Sample Preparation Kit for Illumina (Illumina) following the manufacturer's standard protocols. The genome sequencing of 43 *E. coli* isolates was performed by LabGenomics (Korea) using an Illumina NextSeq platform (Illumina, USA) to generate pair-end sequence reads (2 × 150 bp). The qualified sequence reads were assembled using the CLC Genomics Workbench v10.0.1 (Qiagen, German), and the assembled draft genome sequences were annotated using the NCBI Prokaryotic Genome Annotation Pipeline. The *E. coli* draft genome sequences and their annotation data were deposited in the NCBI GenBank database. The associated Bioproject number is PRJNA889329, and the GenBank accession numbers of the *E. coli* isolates are listed in Table 2.

2.5. *In silico* serotyping

In silico serotyping was performed with the draft genome sequences of 43 *E. coli* isolates using SerotypeFinder 2.0

(https://cge.food.dtu.dk/services/SerotypeFinder/) with the sequence database of *E. coli* O-antigen-related genes (*wzx* and *wzy* variants).

2.6. Antiserum serotyping

The O-serotypes of 43 *E. coli* isolates were determined using *E. coli* O Antisera For Use With Boiled Culture (SSI Diagnostica, Denmark). In addition, 11 *E. coli* reference strains for which the O-serotypes were already known were used to verify the accuracy of the antiserum serotyping. The boiled culture was prepared via the incubation of selected *E. coli* isolates in LB broth for 12 h at 37°C followed by boiling at >90°C for 1 h. After boiling, the culture was left for 1 h before use to allow the precipitation of bacterial debris. An equal volume of antiserum was mixed with 80 µl of the boiled culture solution, and the mixture was dispensed into 96-well round-bottomed microliter plates (SPL, Korea). 1X phosphate-buffered saline was used as the negative control and added to each well of the plate. After 12 h of incubation at 52°C, the presence of agglutination was confirmed and compared with the negative control. A Gel Doc EZ Imager (Bio-Rad) was used to visualize the results.

2.7. Collection and comparative analysis of O-serotype–specific gene sequences

To differentiate the serotypes of *E. coli*, the DNA sequences of O-serotype–specific genes (*wzy* and *wzx* encoding O-antigen polymerase and O-antigen flippase, respectively) were collected from the database (version 1.0.0) of SerotypeFinder 2.0. To determine the O-serotype–specific sequences of *wzy* or *wzx*, the collected *wzy* and *wzx* sequences were compared separately using ClustalW2 (The European Bioinformatics Institute; EMBL-EBI, United Kingdom). In addition, the complete genome sequences of *E. coli* were collected from the NCBI GenBank database, and their O-serotypes were determined using SerotypeFinder 2.0. The O-serotypes identified for these *E. coli* complete genome sequences were then compared with the O-serotype–specific sequences using NCBI BLASTN. After confirming the O-serotype specific sequences, those of *wzy* or *wzx* were used to design the O-serotype specific primer sets and associated probes.

2.8. Oligonucleotide primer and probe design

Based on the O-serotype-specific sequences of the *wzy* and *wzx* genes of different *E. coli* serotypes, PCR primer–probe sets were

designed using NCBI Primer-BLAST (https://www.ncbi.nlm.nih.gov/tools/primer-blast/) with amplicon size set to 50–200 bp and the length of the primer and probe set to 20–30 bp. The optimal melting temperatures of the primers and probes were set to 55°C and 60°C, respectively. The candidate primer–probe sets were evaluated for whether they made primer–dimer structures using Multiple Primer Analyzer (Thermo Fisher Scientific, USA). To confirm the O-serotype specificity of the selected primer–probe sets, their binding identities to the *wzy* or *wzx* genes of other O-serotype *E. coli* strains were validated using NCBI BLASTN. After the primer–probe sets were evaluated, they were chemically synthesized by Bionics. In particular, the probes were labeled with a fluorescent dye, such as HEX (6-hexachlorofluorescein), CY5 (indodicarbocyanine), and FAM (6-carboxyfluorescein), and a quencher dye, such as BHQ-1 or 2 (Black Hole Quencher 1 or 2) and TAMRA (6-carboxy-tetramethylrhodamine). The primer and probe sequences and types of probe-labeled fluorescence and quencher dyes are listed in Table 3.

Table 3. List of primer-probe and content of mixture in multiplex PCR

Group	Serotype	Target gene	Probe (Reporter-Quencher dye)	Amplicon length/bp[a]	Content of primer (F, R each)	Content of probe
Group 1	O145	*wzy*O145	FAM-TAMRA	130	15 pmol	15 pmol
	O157	*wzy* O157	HEX-BHQ-1	108	15 pmol	15 pmol
	O26	*wzy* O26	CY5-BHQ-2	78	15 pmol	15 pmol
Group 2	O103	*wzx* O103	FAM-TAMRA	137	15 pmol	15 pmol
	O111	*wzx* O111	HEX-BHQ-1	90	15 pmol	15 pmol
Group 3	O6	*wzy* O6	FAM-TAMRA	131	12 pmol	12 pmol
	O25	*wzy* O25	HEX-BHQ-1	184	25 pmol	25 pmol
	O78	*wzx* O78	CY5-BHQ-2	73	8 pmol	8 pmol
Group 4	O55	*wzx* O55	FAM-TAMRA	161	10 pmol	10 pmol
	O104	*wzx* O104	HEX-BHQ-1	69	25 pmol	25 pmol
	O124	*wzx* O124	CY5-BHQ-2	128	10 pmol	10 pmol
Group 5	O127	*wzy* O127	HEX-BHQ-1	87	25 pmol	25 pmol
	O115	*wzy* O115	FAM-TAMRA	163	10 pmol	10 pmol
	O36	*wzy* O36	CY5-BHQ-2	197	10 pmol	10 pmol
Group 6	O76	*wzy* O76	CY5-BHQ-2	73	10 pmol	10 pmol
	O121	*wzy* O121	FAM-TAMRA	112	10 pmol	10 pmol
	O168	*wzy* O168	HEX-BHQ-1	197	25 pmol	25 pmol
Group 7	O88	*wzy* O88	HEX-BHQ-1	184	15 pmol	15 pmol
	O51	*wzy* O151	FAM-TAMRA	118	15 pmol	15 pmol
Group 8	O177	*wzy* O177	FAM-TAMRA	189	15 pmol	15 pmol
	O116	*wzy* O116	HEX-BHQ-1	114	15 pmol	15 pmol

[a], Size of amplicon, base-pair

2.9. Conventional PCR for the verification of primer sets

2.9.1. Singleplex PCR

The singleplex PCR reaction mixture contained 1 μl of template DNA (40 ng/μl), 0.75 μl of forward and reverse primers (20 μM each), 2.5 μl of 10X *Taq* buffer, 2.5 μl of dNTP mixture (2 mM each; MGMed), and 0.5 μl of *Taq* DNA polymerase (1 U/μl; MGMed), and the total reaction volume was adjusted to 25 μl using Ultrapure Water (Welgene). The PCR conditions were as follows: 1 cycle of 94°C for 3 min; 35 cycles of 94°C for 30 s, 62°C for 1 min, and 72°C for 1 min; and 1 cycle of 72°C for 5 min. PCR amplicons were separated using standard agarose gel electrophoresis with a 4% agarose gel.

2.9.2. Cross-check PCR

For cross-check PCR, two sets of template DNA mixtures were prepared. A template DNA mixture was prepared as the test mixture using the genomic DNA of 21 O-serotype *E. coli* strains (final concentration: 20 ng/μl each), and another template DNA mixture was prepared as the negative control with the genomic DNA of 20 O-serotype *E. coli* strains without that of the target O-serotype strain. The composition of the PCR mixture and PCR conditions were the same as those given for singleplex PCR.

2.9.3. Multiplex PCR

For multiplex PCR, the eight groups of primer set mixtures (2–3 primer sets per group), predicted PCR amplicon sizes, and added amounts of forward and reverse primers per group reaction are listed in Table 3. The test template DNA mixture in a selected group contained the genomic DNA of only the target strain in the selected group and all the genomic DNA of O-serotype strains in the other groups. However, the genomic DNA of the remaining strains in the selected group was not included in the test template DNA mixture. The positive control template DNA mixture contained the genomic DNA of 21 O-serotype strains. The multiplex PCR reaction mixture contained 1 μl of template DNA (40 ng/μl of each strain), 0.75 μl of forward and reverse primer mixture containing the specific primer amounts shown in Table 3, 2.5 μl of 10X Taq buffer, 2.5 μl of dNTP mixture (2 mM each; MGMed), and 0.5 μl of Taq DNA polymerase (1 U/μl; MGMed), and the total reaction volume was adjusted to 25 μl using Ultrapure Water (Welgene). The multiplex PCR conditions were the same as those given for singleplex PCR.

2.10. Real-time PCR for the verification of primer–probe sets

2.10.1. Singleplex real-time PCR and detection sensitivity test

The CFX96 Touch Deep Well Real-time PCR Detection system (Bio-Rad) was used to conduct real-time PCR. The reaction mixtures contained 1 μl of DNA template (40 ng/μl), 0.75 μl of each primer (20 μM), 10 μl of MG 2X qPCR Master Mix (MGMed), and 0.75 μl of TaqMan probe (20 μM), and the total reaction volume was adjusted to 20 μl using Ultrapure Water (Welgene). The singleplex real-time PCR conditions were as follows: 1 cycle of 95°C for 5 min, 35 cycles of 95°C for 30 s, and 35 cycles of 60°C for 1 min. The types of 5′-fluoroscence and 3′-quencher added to the dual-labeled single-stranded oligonucleotide probe are listed in Table 3.

To determine the detection limits of the primer–probe sets in real-time PCR, *E. coli* genomic DNA samples were serially diluted from 100 ng/μl to 0.1 pg/μl and used as DNA templates for singleplex real-time PCR. Among the serially diluted genome DNA samples, the minimal amount of DNA showing the real-time PCR signal was determined as the detection limit.

2.10.2. Cross-check real-time PCR

Two sets of template DNA mixtures were prepared as described for the cross-check PCR of conventional PCR. The reaction mixture was the same as that used for singleplex real-time PCR, but the template DNA used was one of the two sets of template DNA mixtures, i.e., with or without the genomic DNA of the target O-antigen strain. The cross-check real-time PCR template mixture and conditions were the same as those given for singleplex real-time PCR.

2.10.3. Multiplex real-time PCR and its use for serotyping

According to the primer–probe set mixture, the reactions of eight groups3, as well as the amounts of each primer and probe used per group reaction, are listed in Table 3. Two sets of template DNA (test and negative control) were prepared as described for the multiplex PCR of conventional PCR. The reaction mixtures contained 1 μl of DNA template (40 ng/μl per strain), 0.75 μl of forward and reverse primer mixture containing the specific primer amounts shown in Table 3, 10 μl of MG 2X qPCR Master Mix (MGMed), and 0.75 μl of TaqMan probe containing the specific probe amount shown in Table 3, and the total reaction volume was adjusted to 20 μl using Ultrapure Water (Welgene).

The multiplex real-time PCR conditions were the same to those given for singleplex real-time PCR. Multiplex real-time PCR was used for the O-serotyping of the 43 selected *E. coli* strains shown in Table 2.

2.11. Food application test

Ground beef, purchased at local food market in Gwanak, Seoul, Korea, was used as the food sample in a food application test. The beef was first flattened and then sterilized using ultraviolet C (UVC) at room temperature for 15 min. After UVC sterilization, 21 different O-serotype strains among 54 *E. coli* strains (Tables 1 and 2) were selected, and each strain was inoculated into 5 g of UVC-sterilized ground beef. Three ground beef samples with different CFU numbers (10^7, 10^4, and 10^1 CFUs per sample) were prepared for each *E. coli* strain. These *E. coli*-inoculated ground beef samples were incubated at 37°C for 12 h, after which each sample was placed in a 3M sterile sample bag containing 45 ml of 0.1% sterilized peptone water and homogenized using a BagMixer CC (Interscience) for 30 s. After homogenization, the meat debris was removed using the filter in the bag, and the filtered solution was transferred to a sterile 15 ml Spinwin centrifuge tube (Tarsons, Korea). The total DNA of the filtered solution was extracted and purified using a G-spin™ Genomic DNA Extraction Kit for Bacteria (INtRON).

Subsequently, real-time PCR was conducted using the extracted total DNA as a template to determine whether each primer–probe set was able to detect and identify the O-serotype of the specific *E. coli* strain in the beef sample.

3. Results

3.1. Isolation, identification, and genome features of *E. coli* isolates

Eleven *E. coli* reference strains were obtained from public culture collections, such as ATCC and NCTC, and their O-serotypes are listed in Table 1. In addition, 43 *E. coli* strains were randomly isolated from environmental samples, swine fecal samples, and vegetables. Subsequent 16S rRNA gene sequence analysis of the 43 *E. coli* isolates confirmed that they were all *E. coli*. Whole-genome sequencing and analysis of these isolates showed that their whole genome sequences (WGSs) consisted of 69–675 contigs, 800–1,879 Mb sequence yields with 145–375× coverage, and 4,520–5,532 coding sequences (Table 4). According to WGS analysis, the average genome size of the 43 *E. coli* isolates was ~5 Mb, which is slightly larger than other *E. coli* genomes, likely owing to their draft genome sequences. The WGSs were used as template DNA sequences for *in silico* serotyping and the confirmation of specific O-serotyping PCR primer–probe sets.

Table 4. *E. coli* isolates and their genome sequence information

Strain	Assembly	Contigs	Yield	Coverage	CDSs
SNU-Eco001	Draft	459	1,240,783,006	248.16	5,387
SNU-Eco002	Draft	575	1,291,154,190	258.23	5,113
SNU-Eco003	Draft	392	1,307,149,620	261.43	5,267
SNU-Eco004	Draft	83	1,386,975,468	277.40	4,882
SNU-Eco005	Draft	159	1,249,680,530	249.94	5,177
SNU-Eco006	Draft	575	1,171,060,568	234.21	5,354
SNU-Eco007	Draft	392	1,128,100,162	225.62	5,203
SNU-Eco008	Draft	121	1,005,628,592	201.13	4,658
SNU-Eco009	Draft	163	1,062,084,774	212.42	5,046
SNU-Eco010	Draft	102	1,142,510,696	228.50	4,732
SNU-Eco011	Draft	101	1,599,416,160	319.88	5,013
SNU-Eco012	Draft	265	1,879,253,890	375.85	5,037
SNU-Eco013	Draft	159	1,236,175,324	250.52	5,510
SNU-Eco014	Draft	226	1,273,244,986	254.65	4,710
SNU-Eco015	Draft	129	1,214,086,810	242.82	5,055
SNU-Eco016	Draft	168	1,068,109,070	213.62	5,481
SNU-Eco017	Draft	69	1,070,552,250	214.11	4,688
SNU-Eco018	Draft	90	1,289,914,782	257.98	4,795
SNU-Eco019	Draft	227	1,168,589,906	233.72	5,231
SNU-Eco020	Draft	103	1,309,000,000	261.80	5,034
SNU-Eco021	Draft	184	1,739,000,000	347.80	5,388
SNU-Eco022	Draft	160	1,286,000,000	257.20	4,918
SNU-Eco023	Draft	163	1,241,000,000	248.20	5,191
SNU-Eco024	Draft	121	1,231,000,000	246.20	5,167
SNU-Eco025	Draft	125	1,400,000,000	280.00	5,151
SNU-Eco026	Draft	74	1,373,000,000	274.60	4,520
SNU-Eco027	Draft	137	1,173,000,000	234.60	5,348
SNU-Eco028	Draft	181	1,110,228,104	222.05	5,185
SNU-Eco029	Draft	225	1,278,514,282	255.70	5,203
SNU-Eco030	Draft	224	1,228,455,668	245.69	4,909
SNU-Eco031	Draft	675	1,284,654,244	256.93	5,532
SNU-Eco032	Draft	533	800,050,850	160.01	4,970
SNU-Eco033	Draft	185	938,422,720	187.68	4,865
SNU-Eco034	Draft	234	908,831,552	181.77	4,909
SNU-Eco035	Draft	301	935,452,550	187.09	5,035
SNU-Eco036	Draft	257	1,327,601,060	265.52	5,140
SNU-Eco037	Draft	97	1,240,634,120	248.13	4,847
SNU-Eco038	Draft	145	1,772,556,686	322.28	5,132
SNU-Eco039	Draft	316	960,063,738	174.56	5,311
SNU-Eco040	Draft	675	852,645,660	155.03	5,436
SNU-Eco041	Draft	159	1,126,549,996	204.83	5,131
SNU-Eco042	Draft	327	932,969,506	169.63	5,499
SNU-Eco043	Draft	181	801,447,902	145.72	5,231

3.2. *In silico* serotyping

The WGSs of 43 *E. coli* isolates were used for *in silico* serotyping via SerotypeFinder, with the *wzy* and *wzx* gene sequences compared between the WGSs and SerotypeFinder database. The *in silico* serotyping results are shown in Table 5. Interestingly, only 21 O-serotypes were detected among the 43 *E. coli* isolates with 97.69%–100% sequence identity: O145 (2 strains), O157 (2), O26 (1), O103 (2), O111 (2), O6 (2), O25 (2), O55 (1), O104 (2), O127 (2), O115 (3), O36 (2), O76 (3), O121 (3), O168 (3), O88 (3), O51 (3), O177 (3), and O116 (2). However, as these are *wzy/wzx* sequence–based O-serotype predictions, they were confirmed via additional serotyping, i.e., antiserum serotyping, as follows.

Table 5. Result of *in silico* serotyping

	***In silico* serotyping results**		
Strain	**Gene**	**Identity (%)**	**Predicted serotype**
SNU-Eco001	*wzy*	97.91	**O115**
	wzx	100	
SNU-Eco002	*wzy*	100	**O116**
	wzx	99.57	
SNU-Eco003	*wzy*	100	**O121**
	wzx	100	
SNU-Eco004	*wzy*	100	**O127**
	wzx	100	
SNU-Eco005	*wzy*	100	**O168**
	wzx	100	
SNU-Eco006	*wzy*	99.69	**O177**
	wzx	99.37	
SNU-Eco007	*wzy*	97.69	**O36**
	wzx	98.23	
SNU-Eco008	*wzy*	98.55	**O51**
	wzy	98.4	
SNU-Eco009	*wzy*	100	**O76**
	wzx	99.92	
SNU-Eco010	*wzy*	100	**O88**
	wzx	100	
SNU-Eco011	*wzy*	99.64	**O104**
	wzx	100	
SNU-Eco012	*wzy*	100	**O121**
	wzx	100	
SNU-Eco013	*wzy*	99.91	**O25**
	wzx	100	
SNU-Eco014	*wzx*	100	**O25**
SNU-Eco015	*wzy*	97.69	**O36**
	wzx	98.23	
SNU-Eco016	*wzy*	99.74	**O116**
	wzx	99.86	
SNU-Eco017	*wzy*	100	**O88**
	wzx	100	
SNU-Eco018	*wzy*	100	**O88**
	wzx	100	
SNU-Eco019	*wzy*	99.83	**O145**
	wzx	99.8	
SNU-Eco020	*wzy*	100	**O76**
	wzx	99.92	
SNU-Eco021	*wzy*	100	**O103**
	wzx	100	
SNU-Eco022	*wzx*	100	**O6**
SNU-Eco023	*wzy*	99.91	**O115**
	wzx	100	
SNU-Eco024	*wzy*	100	**O168**
	wzx	100	
SNU-Eco025	*wzy*	99.69	**O177**
	wzx	99.44	

SNU-Eco026	*wzy*	99.91	**O127**
SNU-Eco027	*wzy*	100	**O76**
	wzx	99.92	
SNU-Eco028	*wzy*	99.83	**O145**
	wzx	99.84	
SNU-Eco029	*wzy*	100	**O121**
	wzx	100	
SNU-Eco030	*wzy*	99.64	**O104**
	wzx	100	
SNU-Eco031	*wzx*	100	**O6**
SNU-Eco032	*wzy*	99.84	**O177**
	wzx	99.92	
SNU-Eco033	*wzy*	98.75	**O51**
	wzy	98.46	
SNU-Eco034	*wzy*	98.75	**O51**
SNU-Eco035	*wzy*	100	**O115**
	wzx	100	
SNU-Eco036	*wzx*	100	**O26**
	wzy	99.92	
SNU-Eco037	*wzy*	97.75	**O55**
	wzx	98.59	
SNU-Eco038	*wzy*	100	**O157**
	wzx	100	
SNU-Eco039	*wzy*	100	**O157**
	wzx	100	
SNU-Eco040	*wzx*	100	**O111**
SNU-Eco041	*wzy*	100	**O111**
SNU-Eco042	*wzy*	100	**O103**
	wzx	100	
SNU-Eco043	*wzy*	100	**O168**
	wzx	100	

3.3. Antiserum serotyping

Escherichia coli O-antisera consists of polyclonal antibodies produced by rabbits immunized with boiled cells, and the reaction of O-antisera and boiled culture can identify the serotype of *E. coli* through the production of sediments created by antibody–antigen agglutination reactions. As shown in Fig. 1, a positive reaction creates a “grey carpet” that covers the bottom of the well, whereas a negative reaction is shown by a small white dot in the middle of the well bottom, like in the (−) control. This antiserum serotyping of 11 *E. coli* reference strains and 43 *E. coli* isolates revealed 21 O-serotypes: O145 (NCCP 15954; SNU-Eco019, Eco028), O157 (ATCC 43895; SNU-Eco039, Eco038), O26 (NCCP 15961; SNU-Eco036), O103 (NCCP 15956; SNU-Eco021, Eco042), O111 (ATCC 43887; SNU-Eco040, Eco 041), O6 (NCCP 15732; SNU-Eco022, Eco031), O25 (NCCP 16186; SNU-Eco014, Eco013), O78 (ATCC 35401), O55 (NCCP 13997; SNU-Eco037), O104 (NCCP 15654; SNU-Eco011, Eco030), O124 (ATCC 43893), O127 (SNU-Eco004, Eco026), O115 (SNU-Eco001, Eco023, Eco035), O36 (SNU-Eco007, Eco015), O76 (SNU-Eco009, Eco020, Eco027), O121 (SNU-Eco003, Eco029, Eco012), O168 (SNU-Eco005, Eco036, Eco043), O88 (SNU-Eco010, Eco017, Eco018), O51 (SNU-Eco008,

Eco033, Eco034), O177 (SNU-Eco006, Eco025, Eco032), and O116 (SNU-Eco002, Eco016). As expected, the O-serotypes of all 11 *E. coli* reference strains were the same as the 11 known O-serotypes (Table 1), indicating that the antiserum serotyping conducted in this study is valid. However, the antiserum serotyping of the 43 *E. coli* isolates revealed only 19 O-serotypes. The *E. coli* reference strains did not exhibit O127, O115, O36, O76, O121, O168, O88, O51, O1777, and O116, whereas the *E. coli* isolates did not exhibit O78 and O124. Nevertheless, the antiserum serotyping results of the 43 *E. coli* isolates matched the *in silico* serotyping results exactly, substantiating the accuracy of the *in silico* serotyping performed in this study.

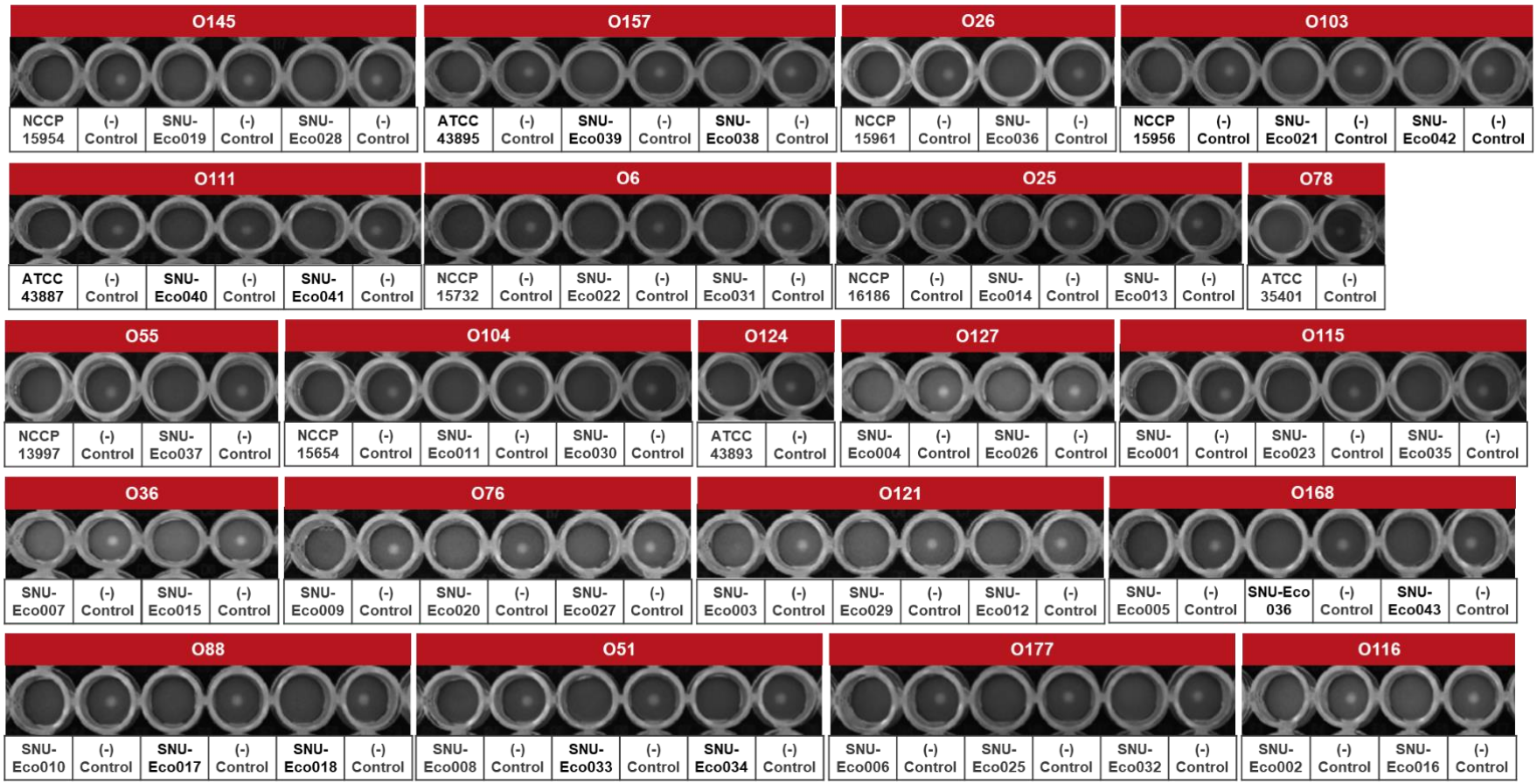


Fig. 1. Result of antiserum serotyping.

There were precipitation reactions in all tested 54 *E. coli* strains. (-) Control is 1X PBS buffer.

3.4. O-serotype grouping and the selection of representative strains

Based on the *in silico* serotyping and antiserum serotyping results, 21 O-serotypes were determined, and 21 strains were selected as O-serotype representative strains: NCCP 15954 (O145), ATCC 43895 (O157), NCCP 15961 (O26), NCCP 15956 (O103), ATCC 43887 (O111), NCCP 15732 (O6), NCCP 16186 (O25), ATCC 35401 (O78), NCCP 13987 (O55), NCCP 15654 (O104), ATCC 43893 (O124), SNU-Eco004 (O127), SNU-Eco001 (O115), SNU-Eco007 (O36), SNU-Eco009 (O76), SNU-Eco003 (O121), SNU-Eco005 (O168), SNU-Eco010 (O88), SNU-Eco008 (O51), SNU-Eco006 (O177), and SNU-Eco002 (O116). The selected representative strains were used to develop and evaluate O-serotype–specific primer–probe sets as well perform real-time PCR. The remaining 33 strains were also used to confirm the real-time PCR results.

3.5. Selection of O-serotype–specific target genes

A previous study showed that *wzy* encoding O-antigen polymerase and *wzx* encoding O-antigen flippase play important roles in the production of O-antigens; thus, *wzy* and *wzx* sequence analyses can be used as standards for the identification of O-serotypes (L. Wang &

Reeves, 1998). Therefore, these genes were comparatively analyzed using SerotypeFinder for O-serotype identification. The *wzy* and *wzx* genes of the 21 selected O-serotype strains exhibited 30.5%–49.1% and 30.2%–54.0% sequence identities, respectively (Fig. 2). Subsequent comparative sequence analysis of either gene could be sufficient to differentiate and identify the specific O-serotypes of *E. coli* strains, consistent with the key genes of SerotypeFinder; therefore, either *wzy* or *wzx* was randomly selected as a target gene for the design of O-serotype–specific primer–probe sets in this study (Table 3).

The O-serotype–specific primer–probe sets were designed using primer-BLAST, and their secondary structures were predicted and evaluated using Multiple Primer Analyzer (data not shown). Based on these results, the O-serotype–specific primer–probe sets were selected (Table 6). To verify the O-serotype specificity of the selected sets, forward and reverse primers targeting specific O-serotypes were used for multiple alignment analysis with the *wzy* or *wzx* genes of 20 other O-serotype *E. coli* strains. In total, 1,339 complete *E. coli* genomes were obtained from NCBI GenBank, and their O-serotypes were identified using SerotypeFinder (data not shown). Among these genomes, 444 complete *E. coli* genomes were associated with the 21 selected O-

serotypes and used to test the O-serotype specificity of the primer sets. Although the sequence identities between the primer set and target genes in the genomes of the same O-serotype strains were 100%, the sequence identities with the target genes of the other 20 O-serotype strains were very low (Table 7), likely owing to low sequence identity among the same genes of the 21 O-serotype strains (Fig. 2). Comparative sequence analysis of the primer sets and target genes indicated that the forward and reverse primers target only specific O-serotypes. Nevertheless, the O-serotype specificity of the primer–probe set was confirmed experimentally using cross-check and multiplex PCR.

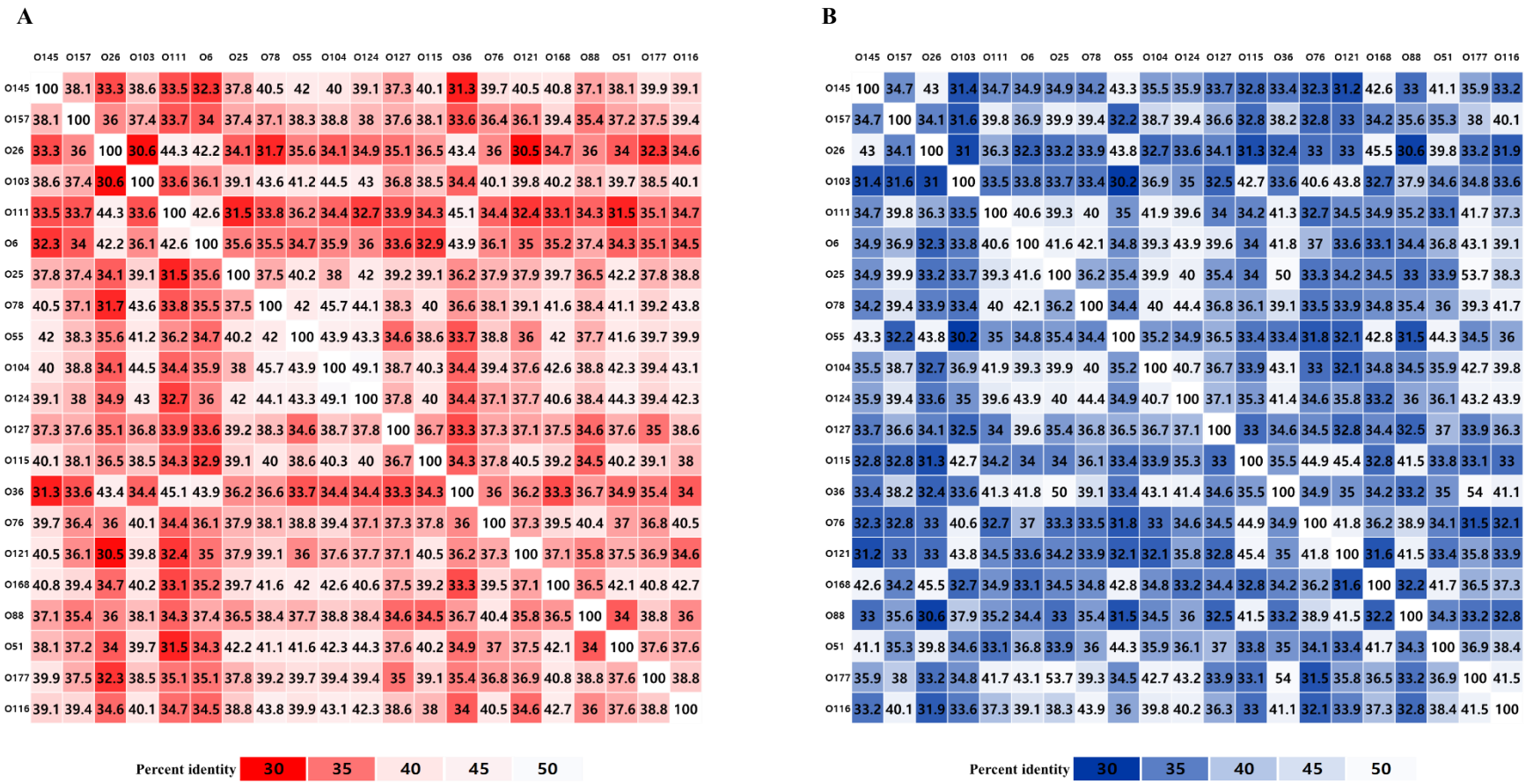


Fig. 2. Comparative sequence analysis of (A) 21 *wzy* genes and (B) 21 *wzx* genes from 21 different *E. coli* O-serotypes. According to comparative analysis, the sequence identities were 30.5%–49.1% and 30.2%–54.0%, respectively.

Table 6. Information of O-serotypes, the sequence of 8 groups of primers and probes for PCR

Group	Serotype	Target gene	Name	Sequence[a] (5' to 3')	Amplicon length/bp[b]
Group 1	O145	*wzy*	O145-F	AGGAGTTATTGTAGGCACGA	130
			O145-R	AGCACCACTAATCGTAGGAT	
			O145-P	FAM-TGGGCTGCCACTGATGGGATTAGGT-TAMRA	
	O157	*wzy*	O157-F	CCATAGCTCGATAAACTGCG	108
			O157-R	ACGGAGAGAAAAGGACCAAA	
			O157-P	HEX-TGGCTGGGAATGCATCGGCCT-BHQ-1	
	O26	*wzy*	O26-F	TAGGCGGTACGCATGAAG	78
			O26-R	TGTATTGTATTAATCGTAGGAATGT	
			O26-P	CY5-TGGCTTGCTGGGTTTATTCCTTGGGA-BHQ-2	
Group 2	O103	*wzx*	O103-F	TTGGAGCGTTAACTGGACCT	137
			O103-R	TGCCGTAGTAACAACAGCCA	
			O103-P	FAM-ACAGCTTGCCAATATCCGGCATCC-TAMRA	
	O111	*wzx*	O111-F	GTCCCTTGCCTTCAACTAGC	90
			O111-R	TGCTTTCATTCTAGCAAACATAGG	
			O111-P	HEX-TGATGTTCACTCTTTCTGCGTCTGCA-BHQ-1	
Group 3	O6	*wzy*	O6-F	ATTGAGCTTGCCGGAATTAA	131
			O6-R	ATCCAATCAGACCAAGTGCT	
			O6-P	FAM-TGGCCTCATACACAGCAAACCCA-TAMRA	
	O25	*wzy*	O25-F	TGGCATACGTCTCCTCTTGT	184
			O25-R	GCAAATGCTGATGGTTCAGG	
			O25-P	HEX-TGTTCGCATTTCTTTGCCTTATAGCA-BHQ-1	
	O78	*wzx*	O78-F	TTGGTTGCAAGTTATTTGATGCC	73
			O78-R	GTGTGAGAGGGATAACATCAACAC	
			O78-P	CY5-CGCTCTCAAAGCGTTTCCATGATG-BHQ-2	

Group 4	O55	*wzx*	O55-F	GGATTGCTAGTTACACAGCTTGA	161
			O55-R	TTCAAGGCCACTGTCAATCG	
			O55-P	FAM-ACGTATTGGCATCTACGTTGGCTCA-TAMRA	
	O104	*wzx*	O104-F	TGTGACCATTGCTTCTCAAGG	69
			O104-R	TCATTTCCGGCTAAACGTGCT	
			O104-P	HEX-TTTCCGGCTAAACGTGC-BHQ-1	
	O124	*wzx*	O124-F	TGCAAGTGTGTTATATTTGGTTGC	128
			O124-R	CCATTTCAAATATCATCATGTATGCGT	
			O124-P	CY5-ACACCATCGCATTAAGTGGCGTC-BHQ-2	
Group 5	O127	*wzy*	O127-F	GGTGTCTCTGCTCAATCAA	87
			O127-R	ACCAAAATAGCCCAACATTAC	
			O127-P	HEX-AGATTGCTTCATCTCCGCTGGGA-BHQ-1	
	O115	*wzy*	O115-F	GAACCCGATTATTTTGTTCGT	163
			O115-R	AGACACGATGCCAATACCAA	
			O115-P	FAM-ACGGGAACAATTGTCGGCGGCT-BHQ-1	
	O36	*wzy*	O36-F	AGTTTTTCAATGGCACAGTC	197
			O36-R	GACCAGTCCAACATATCCAT	
			O36-P	CY5-AGGAGCTGCAGAGTCGGGGCA-BHQ-2	
Group 6	O76	*wzy*	O76-F	AGGATCTTTCGGTTCACAAT	73
			O76-R	GTTGTGAGTATAAGCCCCC	
			O76-P	CY5-TGATATTGATGGGTTCGGTGTCCGT-BHQ-2	
	O121	*wzy*	O121-F	TGCTGGCATAAACTATCTGT	112
			O121-R	TGAGCTAAAATAACCTGCGT	
			O121-P	FAM-TCAACATGTCGGCCATTTAGCGTTGA-BHQ-1	
	O168	*wzy*	O168-F	TGGGGGAGGAATATTTAGTCA	197
			O168-R	ATTTATCGCAGCGTTTCATC	
			O168-P	HEX-TGCTGGTGGAGGACCACGCA-BHQ-1	

Group 7	O88	*wzy*	O88-F	CTGATTGCAGCTCATATGTC	184
			O88-R	CAATTTACGTCCAGACCAGA	
			O88-P	HEX-TGGGTTCCCAATTGTCAGCATCTGGCA-BHQ-1	
	O51	*wzy*	O51-F	TTTACTACATGAGATGACTCAAAT	118
			O51-R	AGCGATGAACACCCAAATAA	
			O51-P	FAM-GGATTGGGTTAGCAAGTGCCATTGG-BHQ-1	
Group 8	O177	*wzy*	O177-F	GGTATTTTGCTATGCCCTCT	189
			O177-R	TGCAGTCAATACCCTTATACA	
			O177-P	FAM-TGGCTCTTAGGATTGATGGGGTCTGGT-BHQ-1	
	O116	*wzy*	O116-F	CATTGTCAGAGAGGGAGC	114
			O116-R	TTCATAAAAGTCCGTAGCGT	
			O116-P	HEX-ACGGTAAAAGAATCCGCGCCCC-BHQ-1	

[a], Probes were labeled with one of fluorescence dyes such as FAM (6-carboxyfluorescein), HEX (6- hexachlorofluorescein), and CY5 (indodicarbocyanine), and one of quencher dyes such as TAMRA (6-carboxy-tetramethylrhodamine), BHQ-1 and BHQ-2 (Black hole quencher-1 and 2)

[b], Size of amplicon, base-pair

Table 7. *In silico* cross-checking result using BLASTN

No.	Serotype	NCBI complete genome No.	Primer sequences (Forward / Reverse)																			
			O145	O157	O26	O103	O111	O6	O25	O78	O55	O104	O124	O127	O55	O36	O76	O25	O168	O88	O51	O177	O116
1	O145	14	14 / 14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	O157	169	0	169 / 169	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	O26	18	0	0	18 / 18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	O103	12	0	0	0	12 / 12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	O111	19	0	0	0	0	19 / 19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	O6	38	0	0	0	0	0	38 / 38	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	O25	83	0	0	0	0	0	0	83 / 83	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	O78	10	0	0	0	0	0	0	0	10 / 10	0	0	0	0	0	0	0	0	0	0	0	0	0
9	O55	9	0	0	0	0	0	0	0	0	9 / 9	0	0	0	0	0	0	0	0	0	0	0	0
10	O104	21	0	0	0	0	0	0	0	0	0	21 / 21	0	0	0	0	0	0	0	0	0	0	0
11	O124	2	0	0	0	0	0	0	0	0	0	0	2 / 2	0	0	0	0	0	0	0	0	0	0
12	O127	5	0	0	0	0	0	0	0	0	0	0	0	5 / 5	0	0	0	0	0	0	0	0	0
13	O115	9	0	0	0	0	0	0	0	0	0	0	0	0	9 / 9	0	0	0	0	0	0	0	0
14	O36	3	0	0	0	0	0	0	0	0	0	0	0	0	0	3 / 3	0	0	0	0	0	0	0
15	O76	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1 / 1	0	0	0	0	0	0
16	O121	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7 / 7	0	0	0	0	0
17	O168	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5 / 5	0	0	0	0
18	O88	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3 / 3	0	0	0
19	O51	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5 / 5	0	0
20	O177	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5 / 5	0
21	O116	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6 / 6
	TOTAL	444																					

3.6. Conventional PCR for the verification of primer sets

3.6.1. Singleplex PCR

To evaluate the specificity of the designed primer sets to the target genes, singleplex PCR was performed using the genome of the target O-serotype strain. All PCR amplicons were shown as single bands with expected sizes in the agarose gel, indicating that the primer sets bound to the target genes and the specific gene regions were amplified correctly (Fig. 3). This result substantiates the specificity of the primer sets to the target genes without the presence of nonspecific binding. Nevertheless, the O-serotype specificity was validated using cross-check PCR as follows.

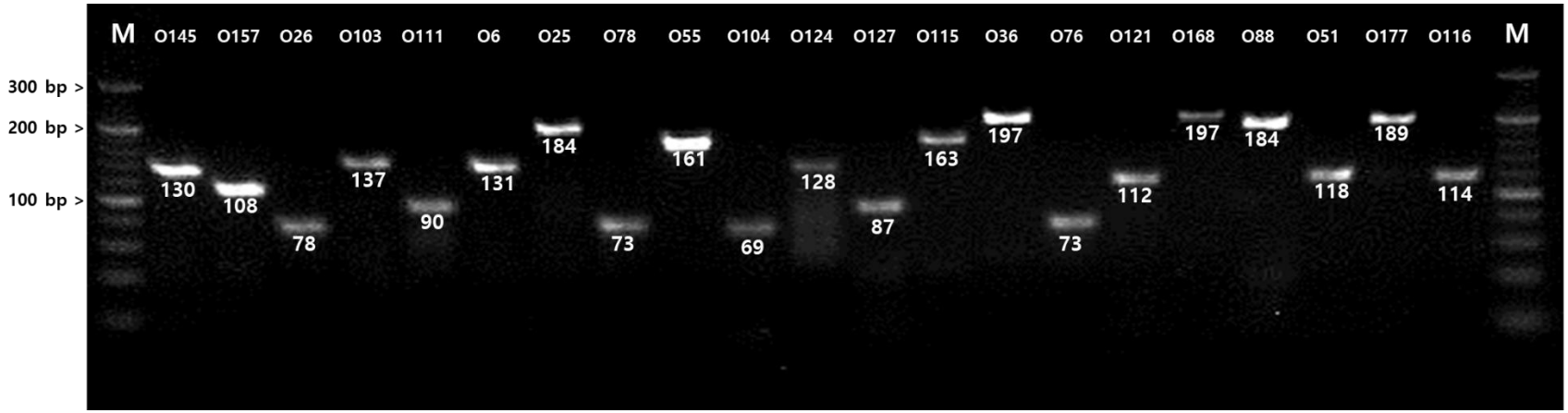


Fig. 3. Conventional singleplex PCR.

The numbers on the gel indicate the base pair (bp) sizes of PCR amplicons, and M indicates O'RangeRuler 20 bp DNA Ladder (Thermo Fisher Scientific).

3.6.2. Cross-check PCR

To verify the O-serotype specificity of each primer set, cross-check PCR was performed with two template DNA mixtures: a test template DNA mixture with the genomic DNA of 21 O-serotype strains and a negative control template DNA mixture with the genomic DNA of 20 O-serotype strains excluding the genomic DNA of the target strain. Cross-check PCR with the specific O-serotype showed a single band only for test template DNA mixture without any nonspecific PCR bands, and the band size was exactly as expected. The same results were obtained for the cross-check PCR of other O-serotype strains, suggesting that only the specific O-serotype strain was targeted and confirming the O-serotype specificity of the primer sets (Fig. 4). However, to increase the efficiency and convenience of O-serotyping via PCR, multiplex PCR must be conducted with multiple validated PCR primer sets as well as multiple genomic DNAs of the target strains.

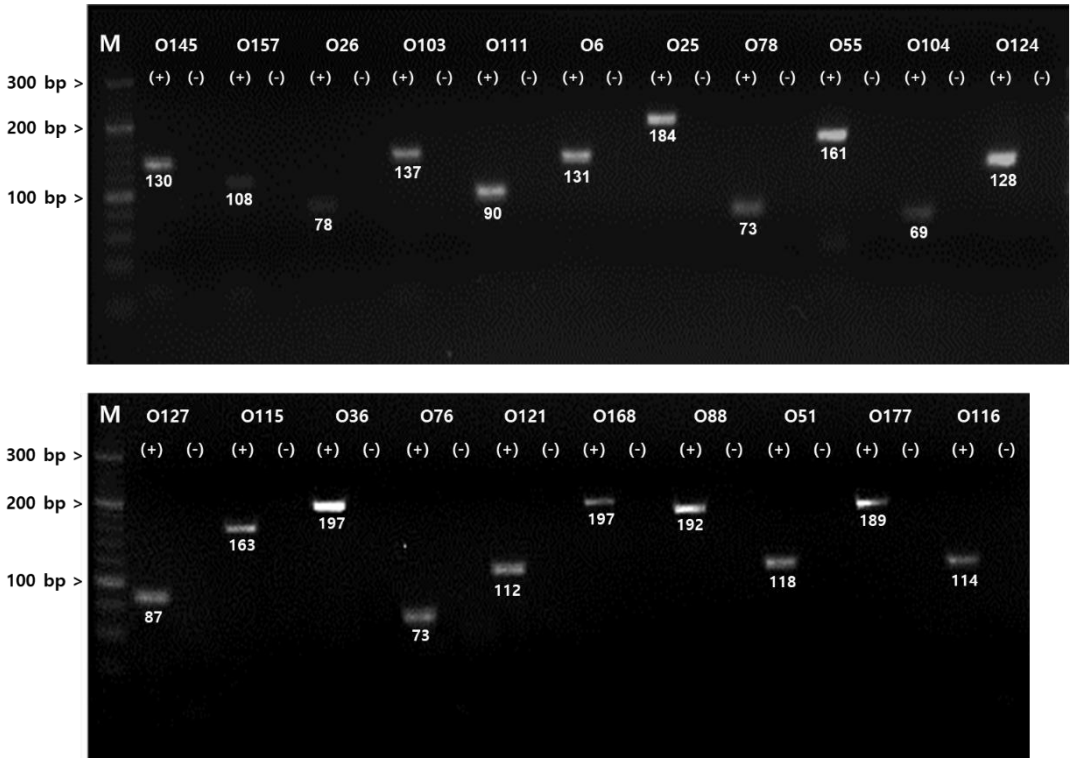


Fig. 4. Conventional cross-check PCR.

(+) indicates that the template DNA mixture contains the genomic DNA of all 21 O-serotype *E. coli* strains, and (−) indicates that the template DNA mixture contains the genomic DNA of 20 O-serotype *E. coli* strains excluding that of the target O-serotype strain. The numbers on the gel and M have the same indications as described for conventional singleplex PCR.

3.6.3. Multiplex PCR

Multiplex PCR was performed using mixtures of the validated PCR primer sets and all target genomic DNA to determine whether O-serotype–specific target genes can be multidetected in one PCR reaction. For multiple O-serotyping with a low number of PCR reactions, the 21 developed O-serotyping primer sets were categorized using Multiple Primer Analyzer (Thermo Fisher Scientific) to determine whether secondary structure formation or interference occurred among the primer sets. These 21 O-serotyping primer sets could be divided into 8 primer–probe combination groups (Groups 1–8; Table 3). Multiplex PCR was performed to confirm these groups, and a single band with a specific size PCR amplicon but without a nonspecific binding band or primer interference was observed for each O-serotype of the selected group (Fig. 5). Therefore, the O-serotypes of 21 different *E. coli* strains were identified in 8 multiplex PCR reactions without any false-positive results. Overall, singleplex, cross-check, and multiple PCRs confirmed that the O-serotyping primer sets functioned perfectly and that the O-serotyping of *E. coli* strains was achieved with high specificity and accuracy using these sets.

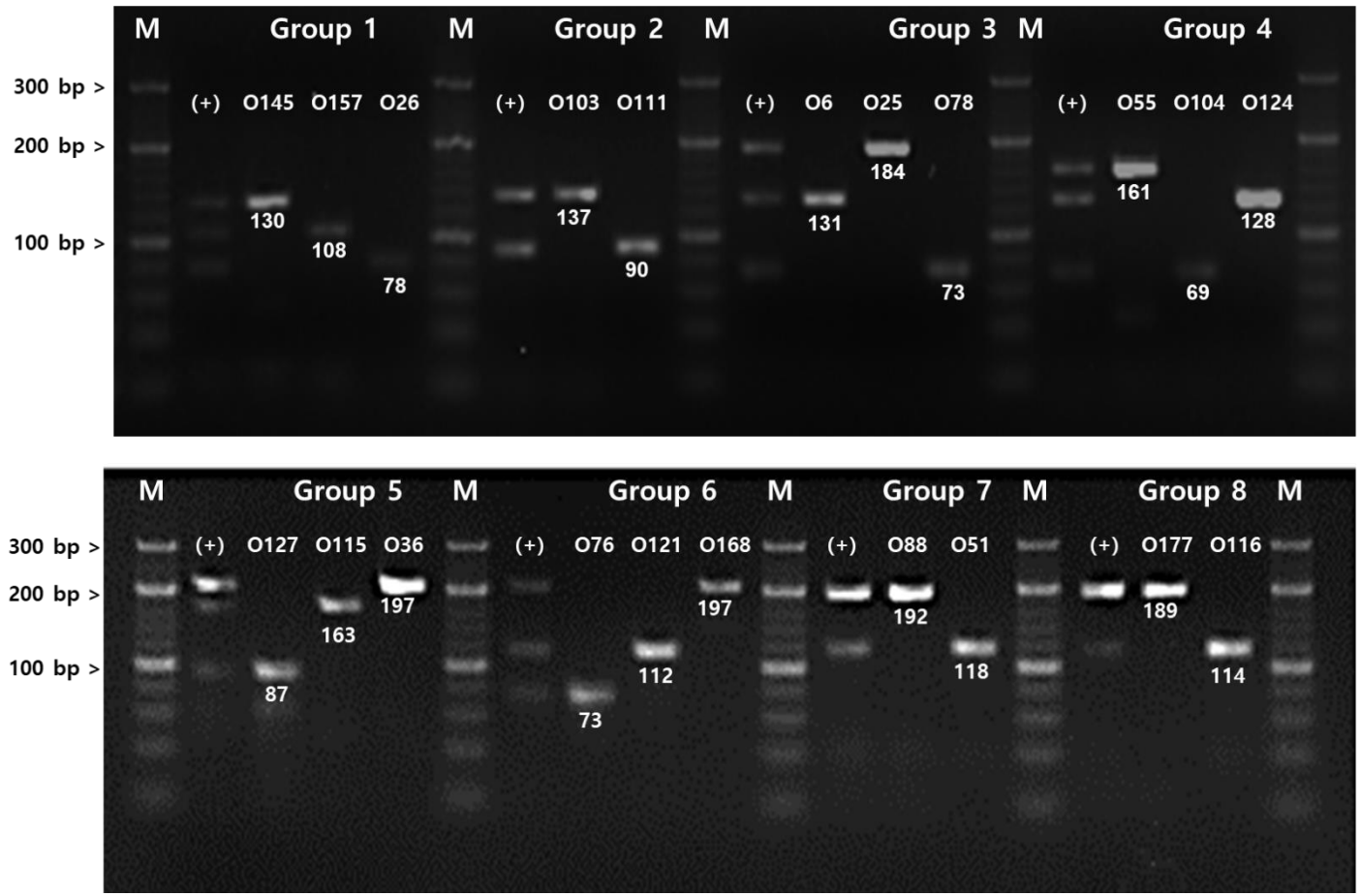


Fig. 5. Conventional multiplex PCR.

(+) indicates that the template DNA mixture contains the genomic DNA of all 21 O-serotype *E. coli* strains. The test template DNA mixture in the selected group contained the genomic DNA of only the target strain in the selected group and all the genomic DNAs of the O-serotype strains in the other groups; however, the genomic DNA of the remaining strains in the selected group was not included in the test template DNA mixture. The numbers on the gel and M have the same indications as described for conventional singleplex PCR.

3.7. Real-time PCR for the verification of primer–probe sets

3.7.1. Singleplex real-time PCR and detection sensitivity test

To achieve the rapid identification of O-serotypes, real-time PCR was used with the validated primer sets and associated probes. First, singleplex real-time PCR was performed and verified using the 21 O-serotype specific *E. coli* strains. Subsequently, the cycle quantification (Cq) values and amplification curves were assessed to evaluate the efficiency and accuracy of O-serotype identification. The Cq value range was 16.06 to 20.82 (Fig. 6), indicating that the O-serotypes of the 21 *E. coli* strains were identified rapidly, i.e., within 21 cycles. According to these Cq values, the specific O-serotype in a selected *E. coli* strain could be identified in ~30–40 min, which is much faster than other identification methods such as antiserum tests, *in silico* serotyping via NGS of genomes, and even general PCR. In all cases, the results of singleplex real-time PCR were clear single curves, indicative of results with high specificity and accuracy. In addition, DNA concentrations of 1–10 pg were detectable via singleplex real-time PCR with the O-serotype specific primer–probe sets, indicating the high sensitivity of the method (Fig. 7). Although the detection limit differed slightly among the

primer–probe sets, the limit would be sufficient to discriminate and identify the O-serotypes of *E. coli* strains under general conditions.

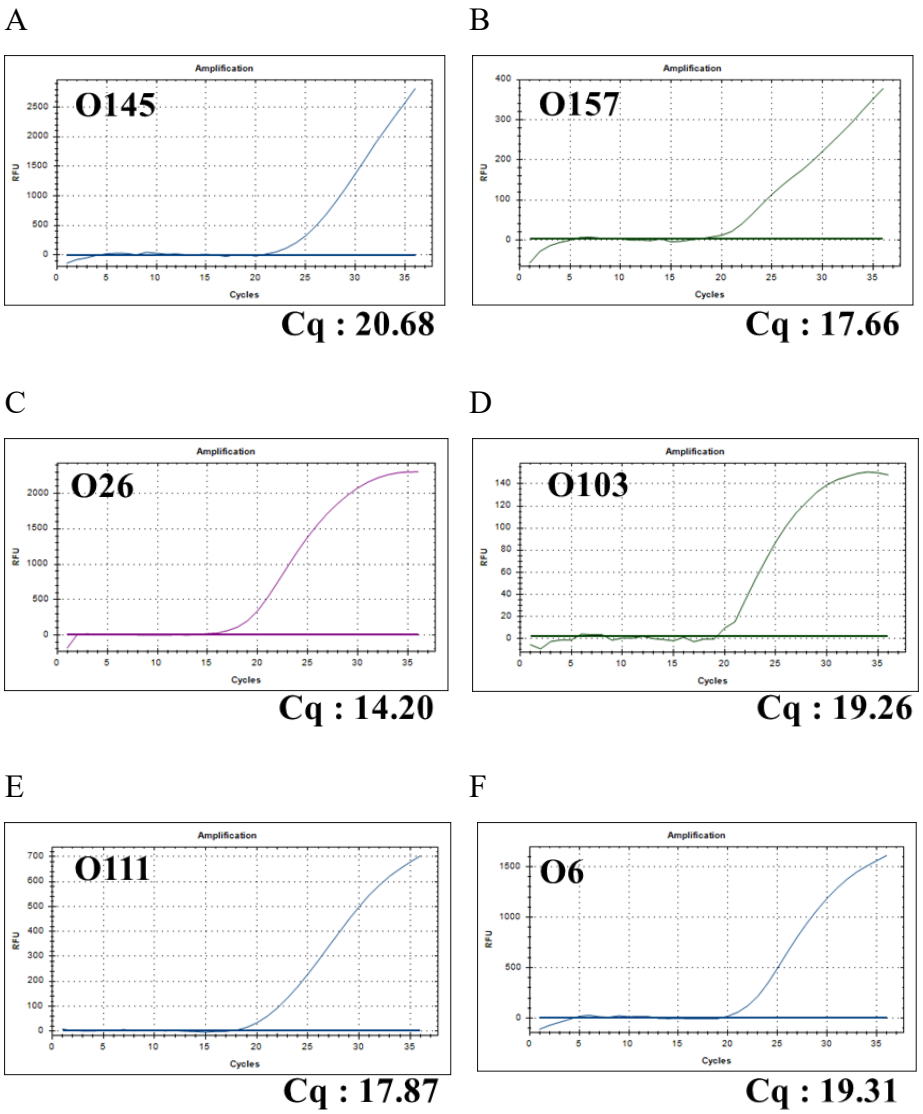
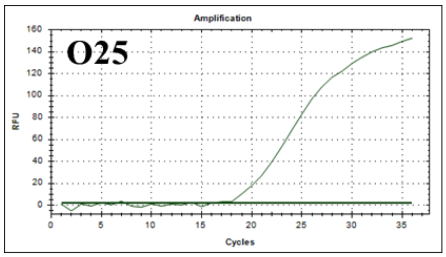


Fig. 6. Amplification curves of Real-time PCR for rapid determination of *E. coli* O-serotype.

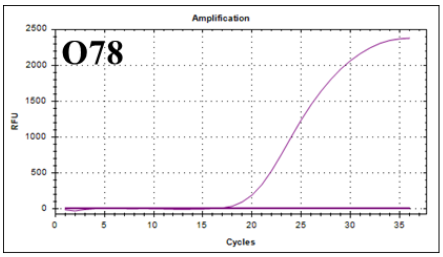
21 specific target DNA were detected in Real-time PCR by TaqMan method. All DNA samples were detected within 20 cycles. (A) O145 , (B) O157, (C) O26, (C) O26, (D) O103, (E) O111, (F) O6

G



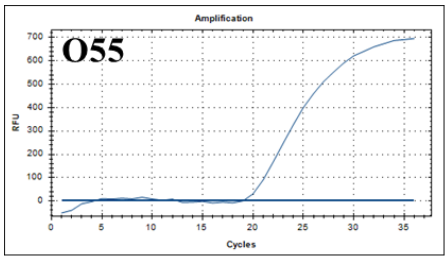
Cq : 16.29

H



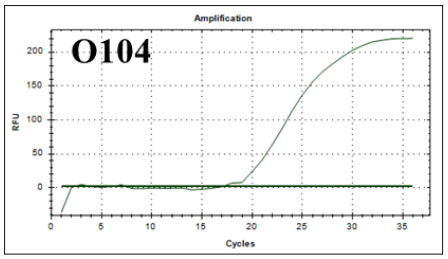
Cq : 16.43

I



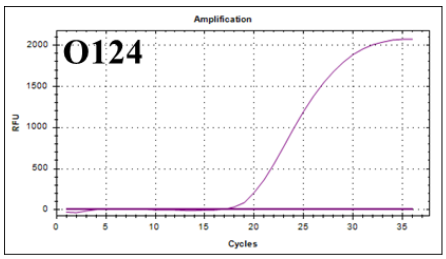
Cq : 19.10

J



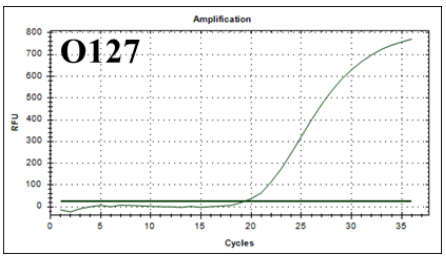
Cq : 17.16

K



Cq : 17.13

L



Cq : 19.37

Fig. 6. Cont. Amplification curves of Real-time PCR for rapid determination of *E. coli* O-serotype.

(G) O25, (H) O78, (I) O55, (J) O104, (K) O124, (L) O127

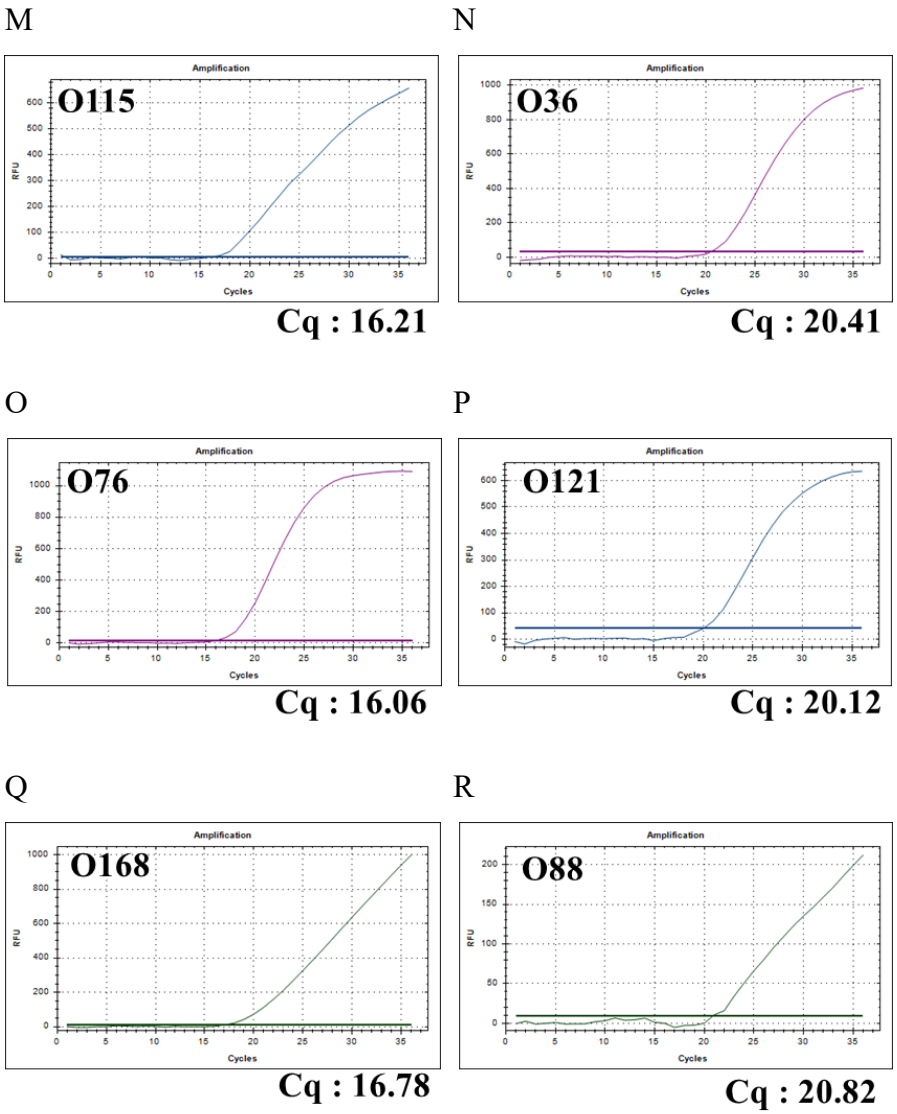
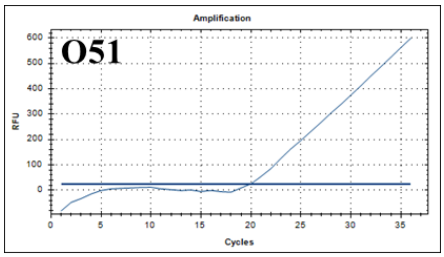


Fig. 6. Cont. Amplification curves of Real-time PCR for rapid determination of *E. coli* O-serotype.

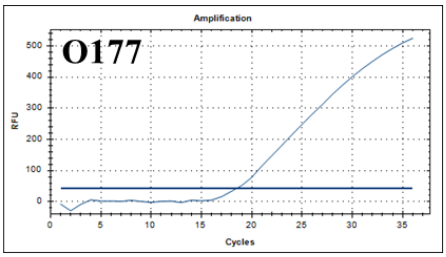
(M) O115, (N) O36, (O) O76, (P) O121, (Q) O168, (R) O88

S



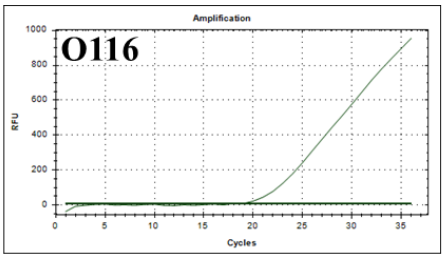
Cq : 20.02

T



Cq : 18.54

U



Cq : 17.86

Fig. 6. Cont. Amplification curves of Real-time PCR for rapid determination of *E. coli* O-serotype.
(S) O51, (T) O177, (U) O116

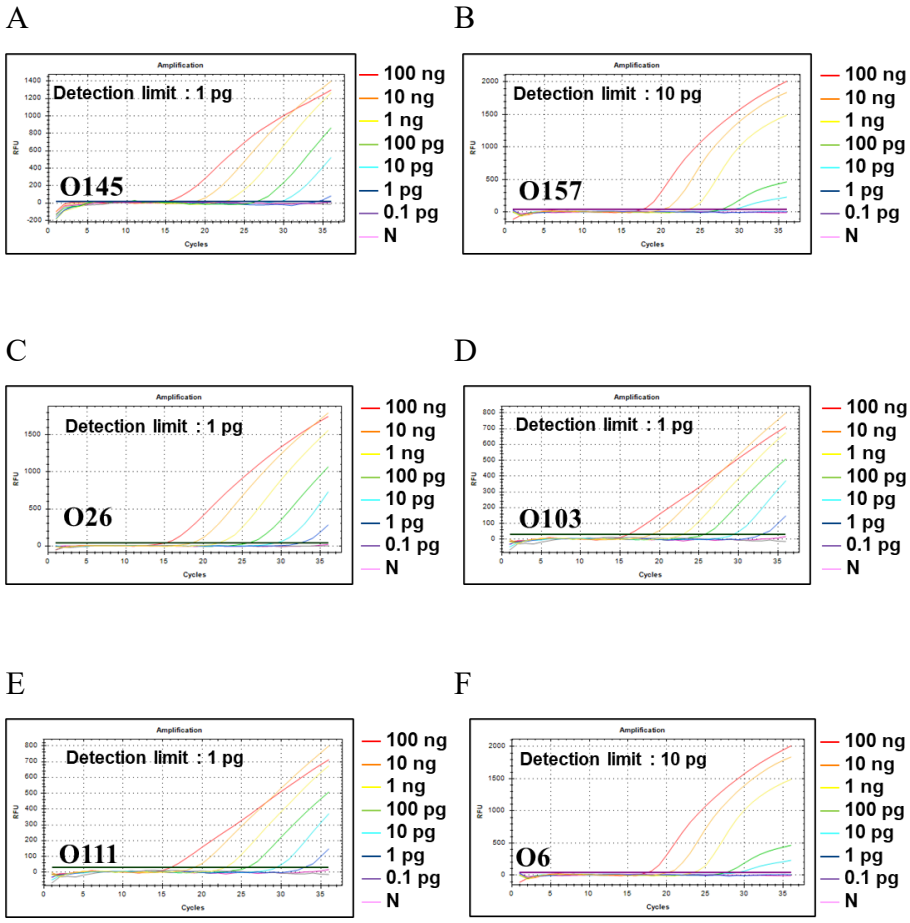


Fig. 7. Amplification curves of Real-time PCR with serially diluted DNA of *E. coli* strains of 21 serotypes.

The DNA samples were diluted 1/10 each 7 times, and results confirmed that the primer-probe sets can detect very low concentration of 1 pg to 10 pg of DNA.. (A) O145, (B) O157, (C) O26, (C) O26, (D) O103, (E) O111, (F) O6

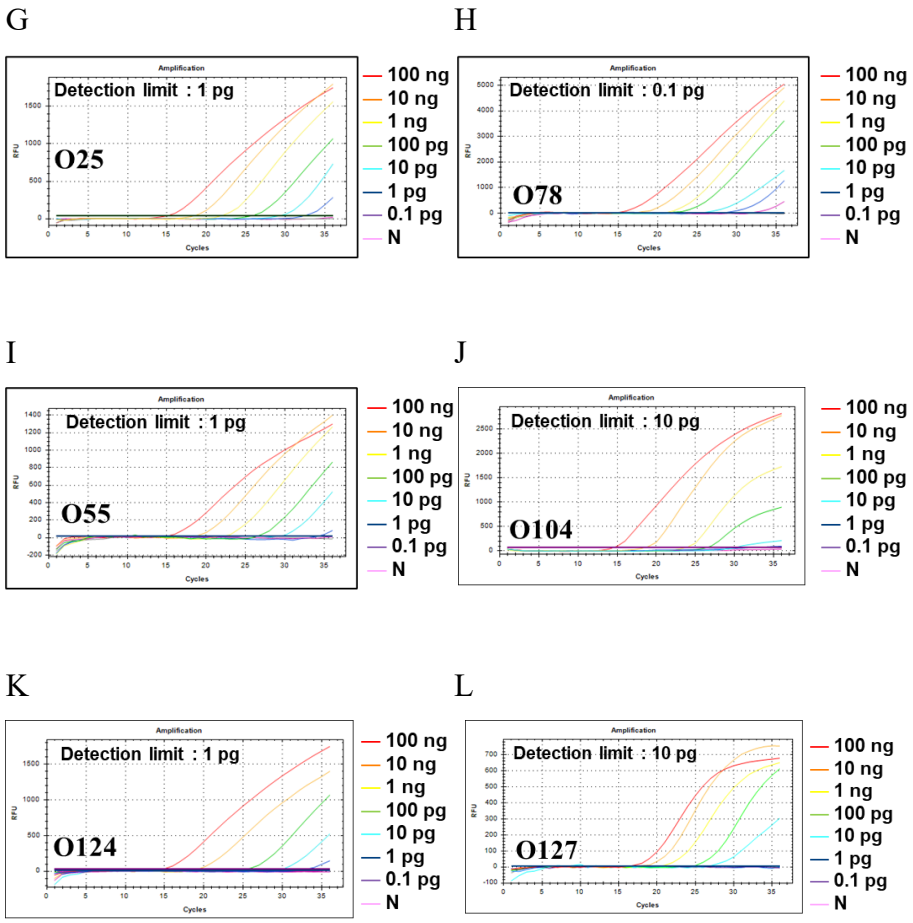
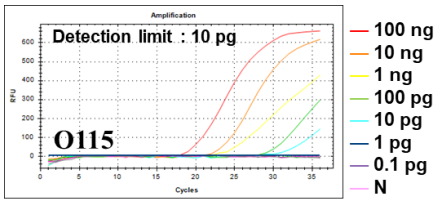


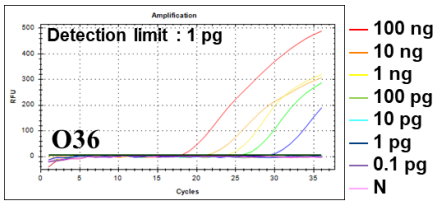
Fig. 7. Cont. Amplification curves of Real-time PCR with serially diluted DNA of *E. coli* strains of 21 serotypes.

(G) O25, (H) O78, (I) O55, (J) O104, (K) O124, (L) O127

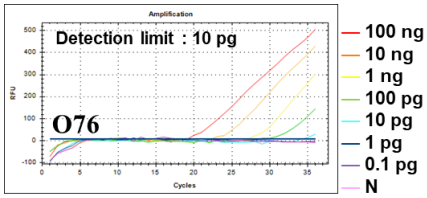
M



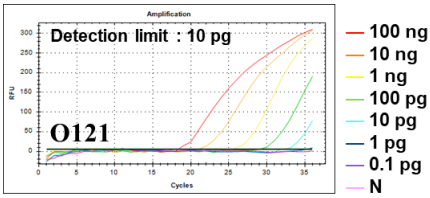
N



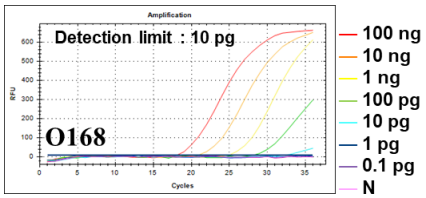
O



P



Q



R

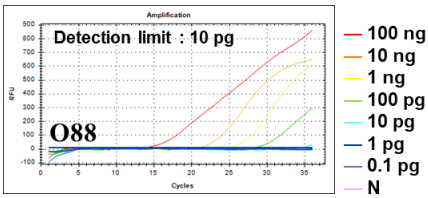
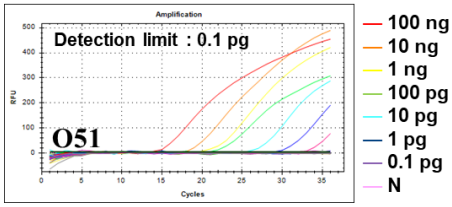


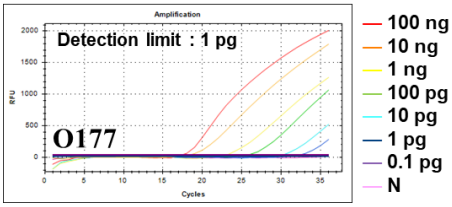
Fig. 7. Cont. Amplification curves of Real-time PCR with serially diluted DNA of *E. coli* strains of 21 serotypes.

(M) O115, (N) O36, (O) O76, (P) O121, (Q) O168, (R) O88

S



T



U

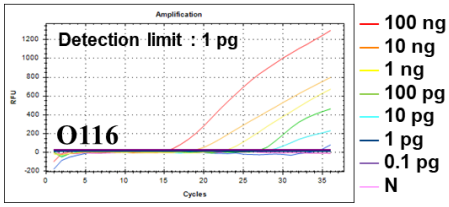


Fig. 7. Cont. Amplification curves of Real-time PCR with serially diluted DNA of *E. coli* strains of 21 serotypes.
(S) O51, (T) O177, (U) O116

3.7.2 Cross-check real-time PCR

To validate the singleplex real-time PCR results, two sets of template DNA mixtures were prepared as described in section 2.10.2, and cross-check real-time PCR was performed. The cross-check real-time PCR results for the negative control template DNA mixture did not show an amplification signal, whereas those for the test template DNA mixture (containing the genome of the target O-serotype *E. coli* strain and those of the other 20 O-serotype strains) showed a clear single amplification curve; thus, the primer–probe sets provided high specificity to the target O-serotype *E. coli* strain without nonspecific amplification (Fig. 8). In summary, the developed real-time PCR method can be used to detect and identify a specific O-serotype strain even in a mixture of various O-antigen *E. coli* strains.

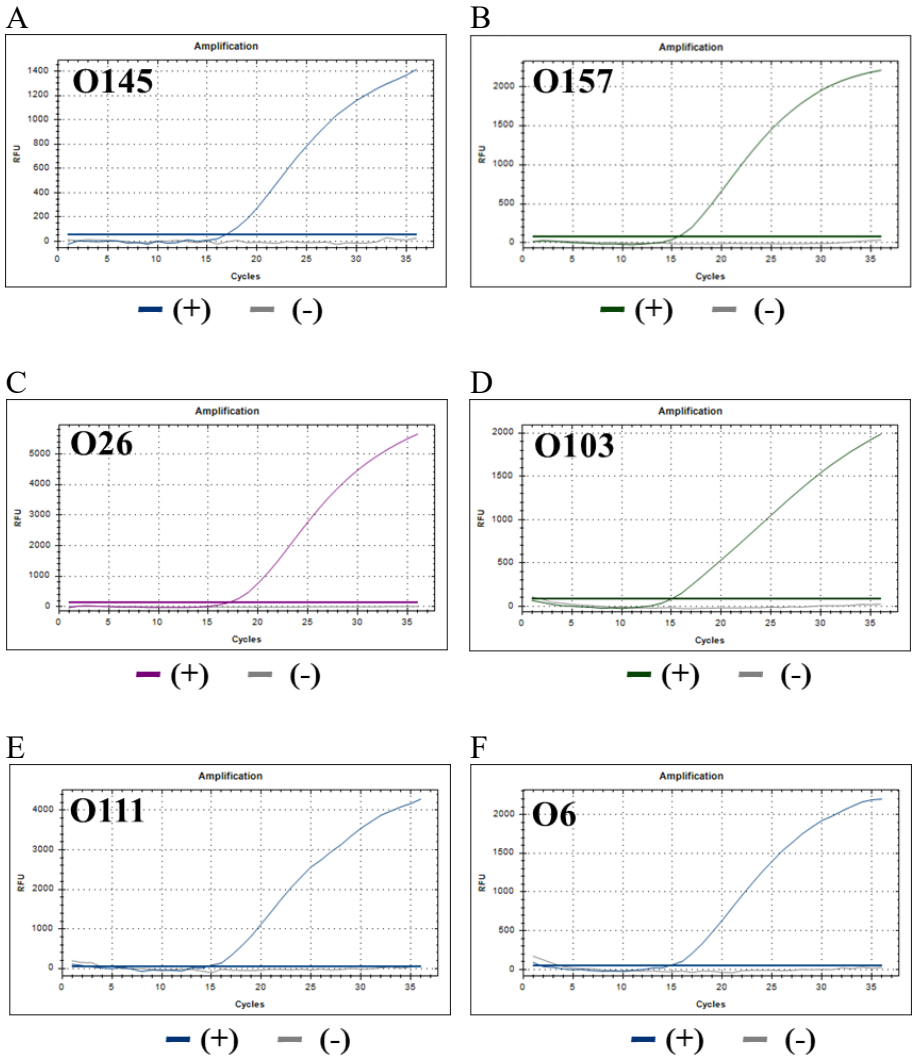


Fig. 8. Amplification curves of cross-check Real-time PCR.

(+), Confirmation of specificity of primer sets in a mixture of chromosomal DNA from E. coli strains of 21 serotypes. (-), Negative control containing DNA of E. coli strains of twenty serotypes except target strain was not amplified. (A) O145, (B) O157, (C) O26, (C) O26, (D) O103, (E) O111, (F) O6

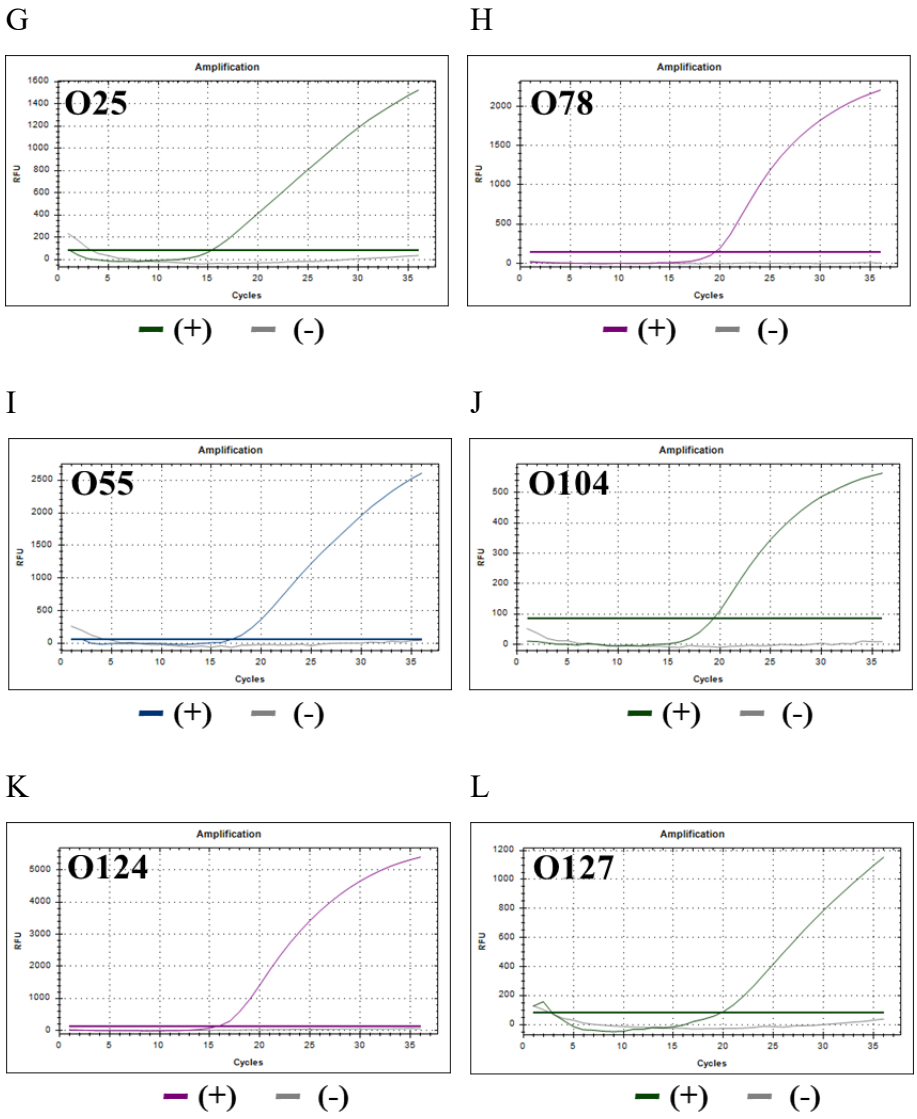


Fig. 8. Cont. Amplification curves of cross-check Real-time PCR.
(G) O25, (H) O78, (I) O55, (J) O104, (K) O124, (L) O127

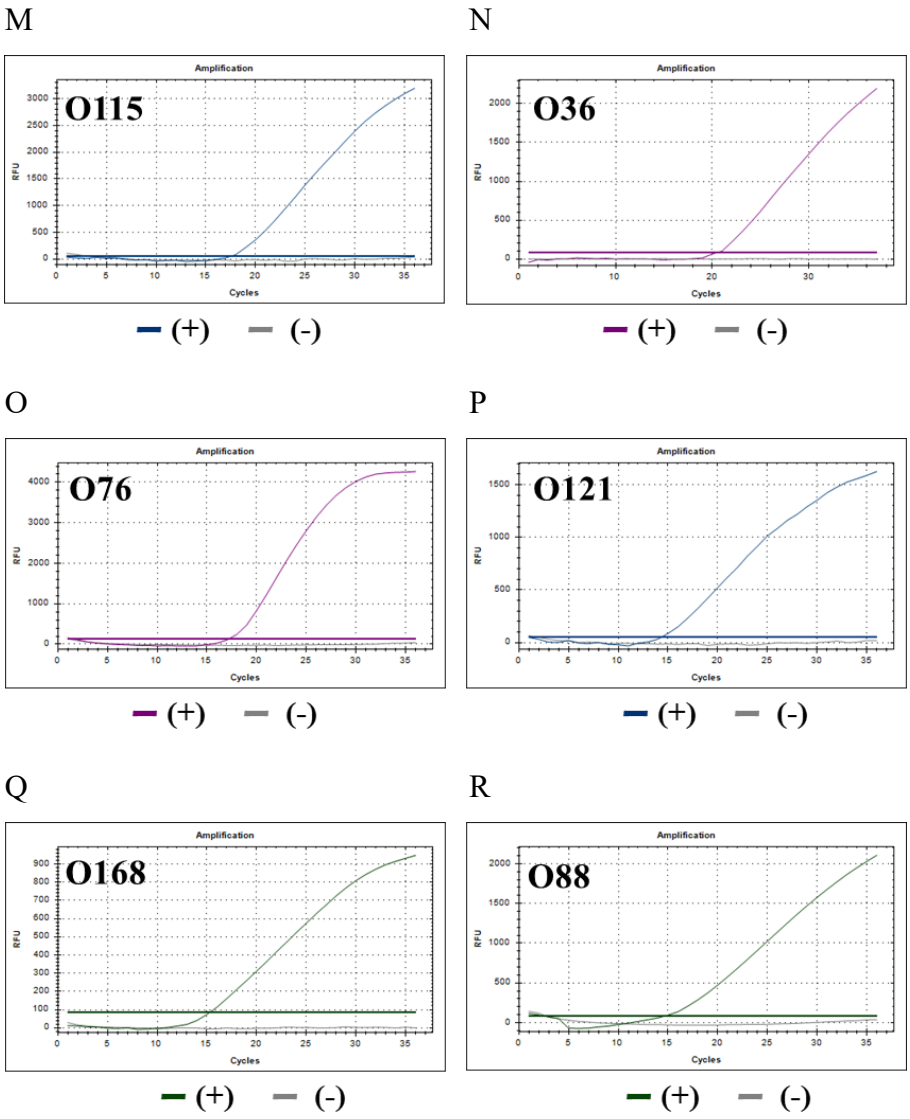
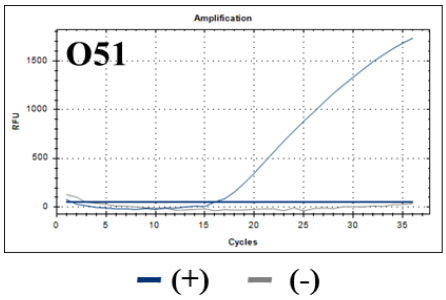
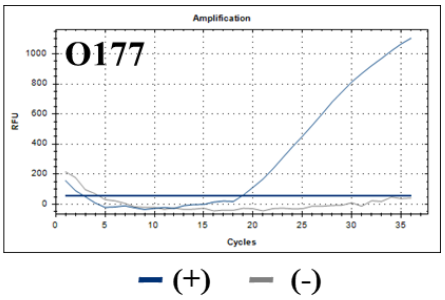


Fig. 8. Cont. Amplification curves of cross-check Real-time PCR. (M) O115, (N) O36, (O) O76, (P) O121, (Q) O168, (R) O88

S



T



U

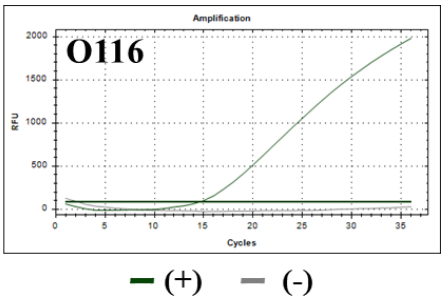


Fig. 8. Cont. Amplification curves of cross-check Real-time PCR. (S) O51, (T) O177, (U) O116

3.7.3. Multiplex real-time PCR and its use for serotyping

Multiplex real-time PCR reduces the cost for O-serotyping by reducing the number of required reactions; however, it requires real-time PCR groups with a mixture of primer–probe sets but no primer–probe interference or secondary primer–probe structure formation. Therefore, the eight groups previously determined using multiplex PCR were also used for multiplex real-time PCR. Additionally, to detect and identify specific O-serotypes simultaneously in one reaction, different fluorescent and quencher dyes were used in the probes of the same group. According to the multiplex real-time PCR results, a single signal curve was obtained for each specific O-serotype in a group without any interference in terms of the detection and identification of the O-serotype, indicating the high sensitivity and specificity of the primer–probe sets as well as the high differentiation of the fluorescent and quencher dyes (Fig. 9). The time required to performed serotyping via singleplex and multiplex real-time PCR was not markedly different. The Cq values of 11 O-serotypes increased slightly when using multiplex real-time PCR, whereas those of the remaining 10 O-serotypes decreased (Table 8). The average Cq value for O-serotypes was slightly reduced when using multiplex real-time PCR (−0.128), suggesting that the method does not

markedly affect the O-serotyping identification time relative to that achieved using singleplex real-time PCR. Furthermore, the sensitivity and specificity of multiplex and singleplex real-time PCRs were similar (Fig. 9).

To validate the multiplex real-time PCR method, O-serotyping of the remaining 33 strains (i.e., those strains remaining from the 54 *E. coli* isolates not chosen as representative strains) was performed (Fig. 10). Interestingly, the multiplex real-time PCR results exactly matched the previous *in silico* serotyping and antiserum analysis results (Table 9), confirming the accurate O-serotyping performance of the developed multiplex real-time PCR method.

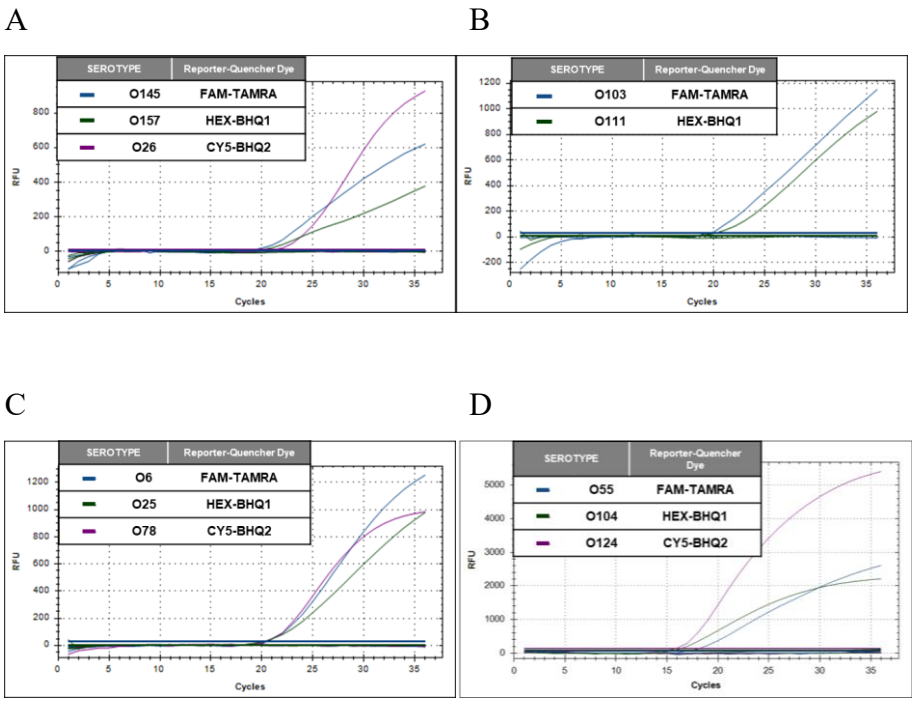
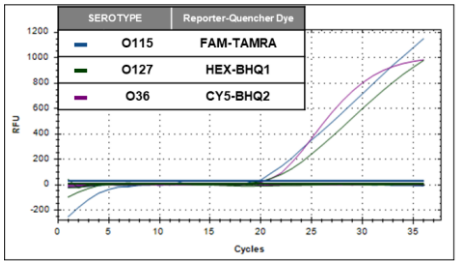


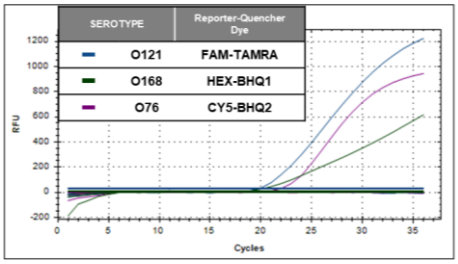
Fig. 9. Amplification curves of multiplex real-time PCR.

The serotype confirmation of 21 O-serotype representative strains (11 O-serotype reference strains and 10 environment isolates with known O-serotypes following antiserum and *in silico* serotyping) in 8 groups.(A) Group 1: serotyping of O145, O157, and O26. (B) Group 2: serotyping of O103 and O111. (C) Group 3: serotyping of O6, O25, and O78. (D) Group 4: serotyping of O55, O104, and O124.

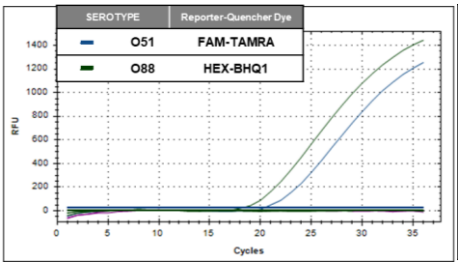
E



F



G



H

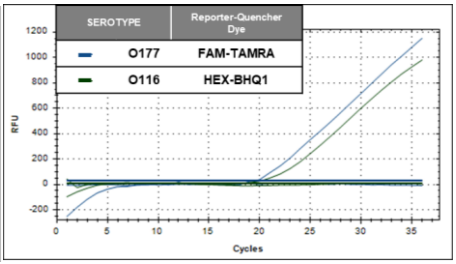


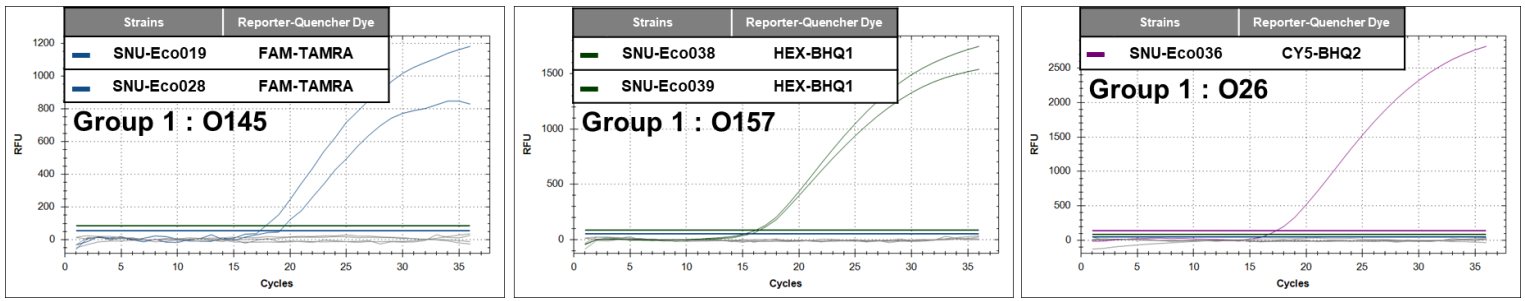
Fig. 9. **Cont.** Amplification curves of multiplex real-time PCR.

(E) Group 5: serotyping of O115, O127, and O36. (F) Group 6: serotyping of O121, O168, and O76. (G) Group 7: serotyping of O51 and O88. (H) Group 8: serotyping of O177 and O116.

Table 8. Cycle quantification (Cq) values for Singleplex Real-time PCR and Cq values for Multiplex Real-time PCR

Serotype	Singleplex Real-time PCR (Cq)[a]	Multiplex Real-time PCR (Cq)[b]	b-a
O145	18.04	19.82	1.78
O157	18.06	20.89	2.83
O26	18.1	20.72	2.62
O103	17.7	12.87	-4.83
O111	19.32	17.35	-1.97
O6	17.98	15.98	-2
O25	17.75	12.57	-5.18
O78	17.83	15.14	-2.69
O55	17.84	16.9	-0.94
O104	19.16	15.69	-3.47
O124	18.15	15.87	-2.28
O127	19.37	20.58	1.21
O115	16.21	20.14	3.93
O36	20.41	19.98	-0.43
O76	16.06	21.45	5.39
O121	20.12	20.32	0.2
O168	16.78	20.54	3.76
O88	20.82	15.45	-5.37
O51	20.02	20.68	0.66
O177	18.54	20.06	1.52
O116	17.86	20.43	2.57
Average			-0.1281

A



B

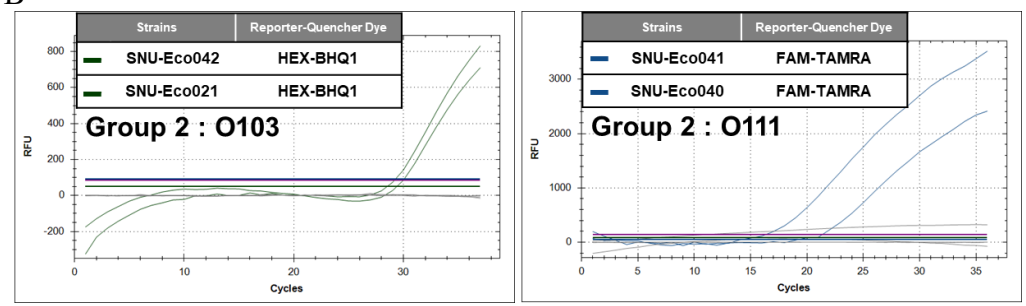
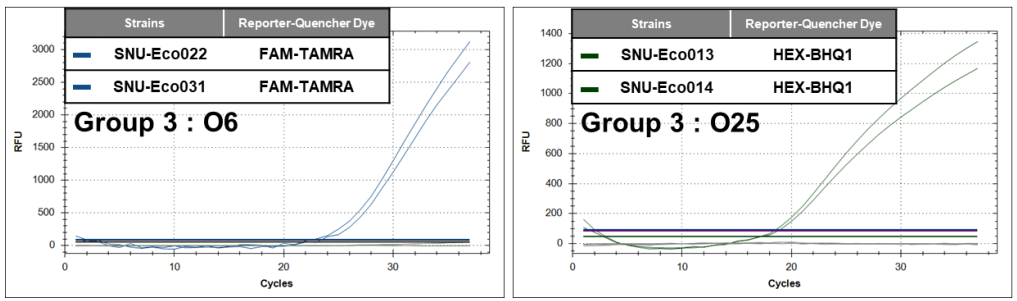


Fig. 10. Additional multiplex real-time PCR for the serotype determination of the remaining 33 *E. coli* environment isolates.

(A) Group 1: serotyping of O145, O157, and O26. (B) Group 2: serotyping of O103 and O111.

C



D

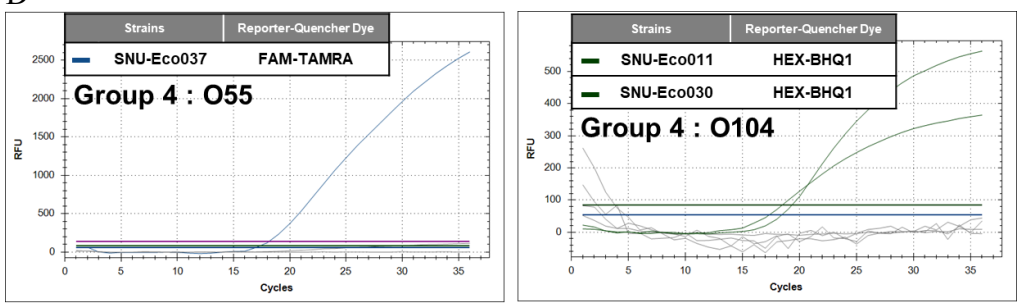
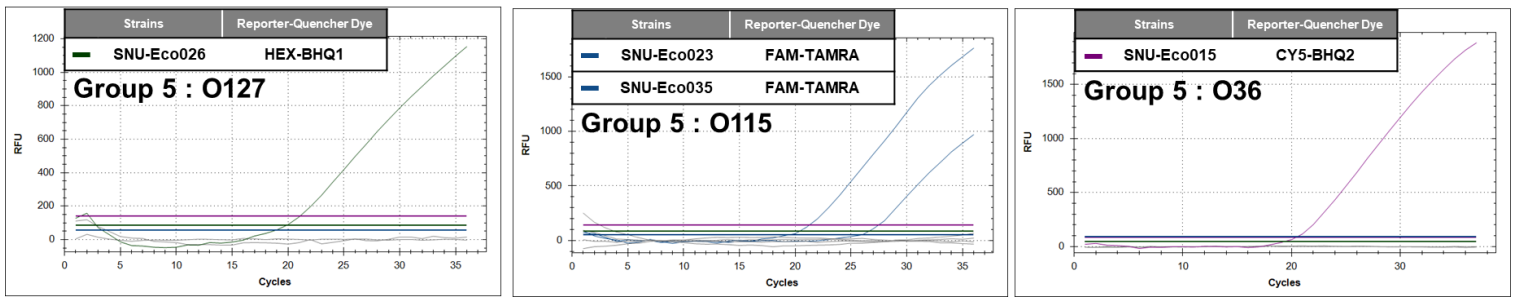


Fig. 10. **Cont.** Additional multiplex real-time PCR for the serotype determination of the remaining 33 *E. coli* environment isolates.

(C) Group 3: serotyping of O6 and O25. (D) Group 4: serotyping of O55 and O104.

E



F

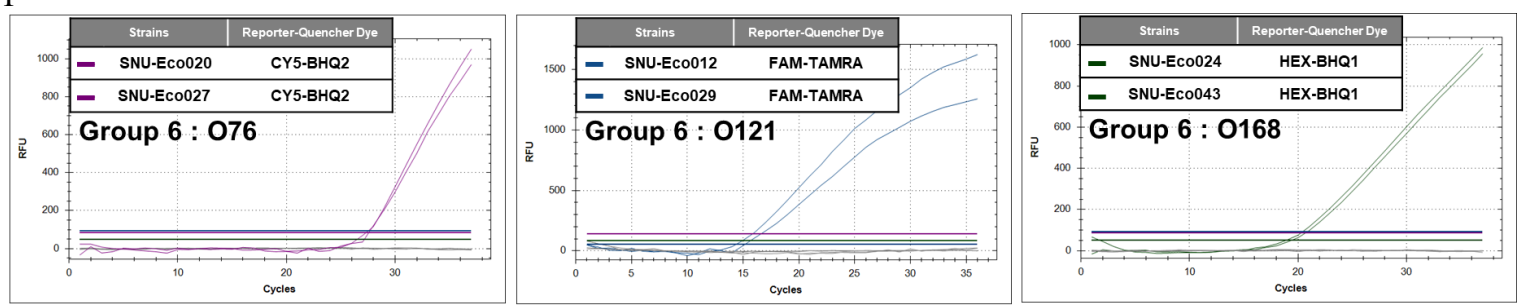
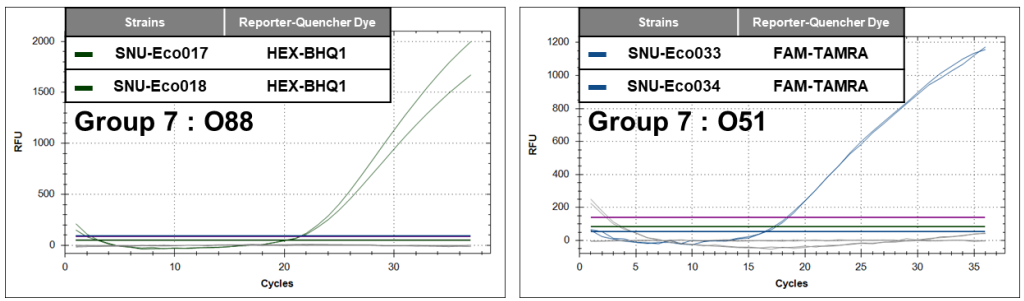


Fig. 10. **Cont.** Additional multiplex real-time PCR for the serotype determination of the remaining 33 *E. coli* environment isolates.

(E) Group 5: serotyping of O115, O127, and O36. (F) Group 6: serotyping of O121, O168, and O76.

G



H

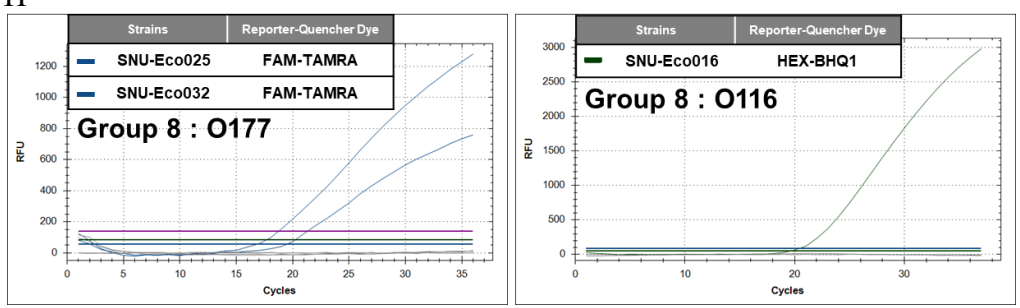


Fig. 10. **Cont.** Additional multiplex real-time PCR for the serotype determination of the remaining 33 *E. coli* environment isolates.

(G) Group 7: serotyping of O51 and O88. (H) Group 8: serotyping of O177 and O116.

Table 9. Comparison of the results of the new PCR serotyping method and existing serotyping methods

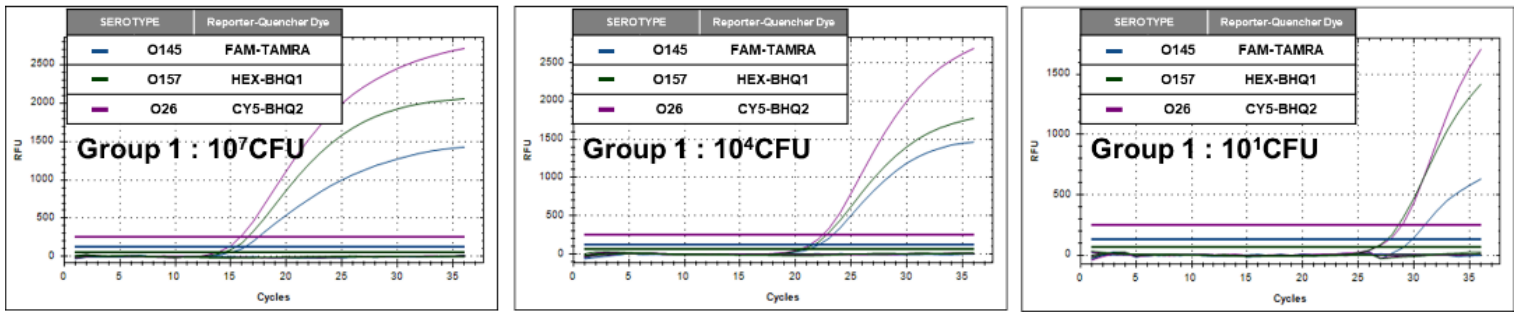
	***In silico* serotyping results**	**Antiserum serotyping results**	**PCR serotyping results**
Strain	**Predicted serotype**	**Predicted serotype**	**Predicted serotype**
SNU-Eco001	O115	O115	O115
SNU-Eco002	O116	O116	O116
SNU-Eco003	O121	O121	O121
SNU-Eco004	O127	O127	O127
SNU-Eco005	O168	O168	O168
SNU-Eco006	O177	O177	O177
SNU-Eco007	O36	O36	O36
SNU-Eco008	O51	O51	O51
SNU-Eco009	O76	O76	O76
SNU-Eco010	O88	O88	O88
SNU-Eco011	O104	O104	O104
SNU-Eco012	O121	O121	O121
SNU-Eco013	O25	O25	O25
SNU-Eco014	O25	O25	O25
SNU-Eco015	O36	O36	O36
SNU-Eco016	O116	O116	O116
SNU-Eco017	O88	O88	O88
SNU-Eco018	O88	O88	O88
SNU-Eco019	O145	O145	O145
SNU-Eco020	O76	O76	O76
SNU-Eco021	O103	O103	O103
SNU-Eco022	O6	O6	O6
SNU-Eco023	O115	O115	O115
SNU-Eco024	O168	O168	O168
SNU-Eco025	O177	O177	O177
SNU-Eco026	O127	O127	O127
SNU-Eco027	O76	O76	O76
SNU-Eco028	O145	O145	O145
SNU-Eco029	O121	O121	O121
SNU-Eco030	O104	O104	O104
SNU-Eco031	O6	O6	O6
SNU-Eco032	O177	O177	O177
SNU-Eco033	O51	O51	O51
SNU-Eco034	O51	O51	O51
SNU-Eco035	O115	O115	O115
SNU-Eco036	O26	O26	O26
SNU-Eco037	O55	O55	O55
SNU-Eco038	O157	O157	O157
SNU-Eco039	O157	O157	O157

SNU-Eco040	O111	O111	O111
SNU-Eco041	O111	O111	O111
SNU-Eco042	O103	O103	O103
SNU-Eco043	O168	O168	O168

3.8. Food application test

Food samples typically include various PCR inhibitors (Al-Soud & Rådström, 2000). In particular, ground beef samples contain potent PCR inhibitors, such as protein, fat, and heme-containing agents; however, such food samples have been widely reported as the sources of *E. coli* foodborne outbreaks (Rangel, Sparling, Crowe, Griffin, & Swerdlow, 2005). Therefore, the multiplex real-time PCR method was used for the *E. coli* O-serotyping of ground beef samples prepared with 10^7, 10^4, and 10^1 CFUs of each O-serotype strain among the 21 O-serotype representative strains. Three single signal curves were shown for each O-serotype of each group, and the Cq value was low when the CFU number was high; thus, *E. coli* O-serotyping was possible in food samples, and the required detection/identification time was inversely proportional to the number of CFUs in the sample (Fig. 11).

A



B

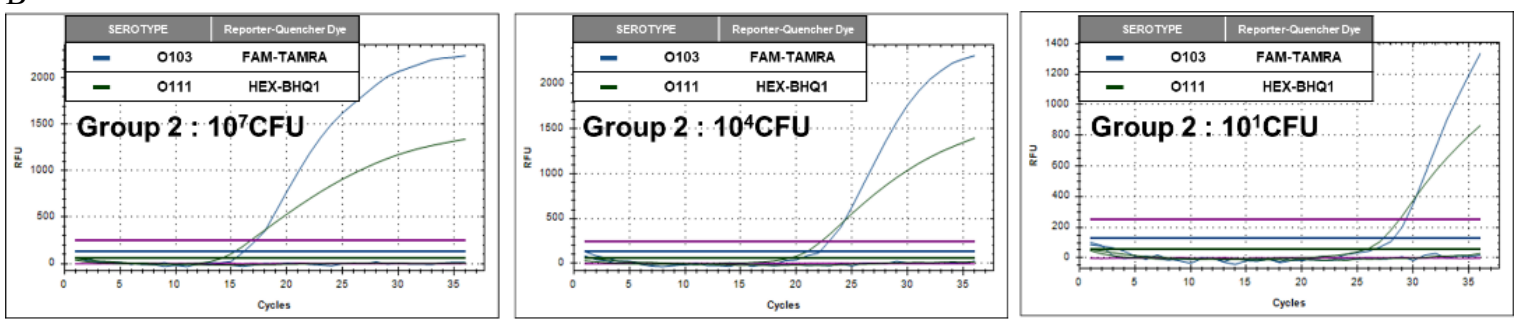
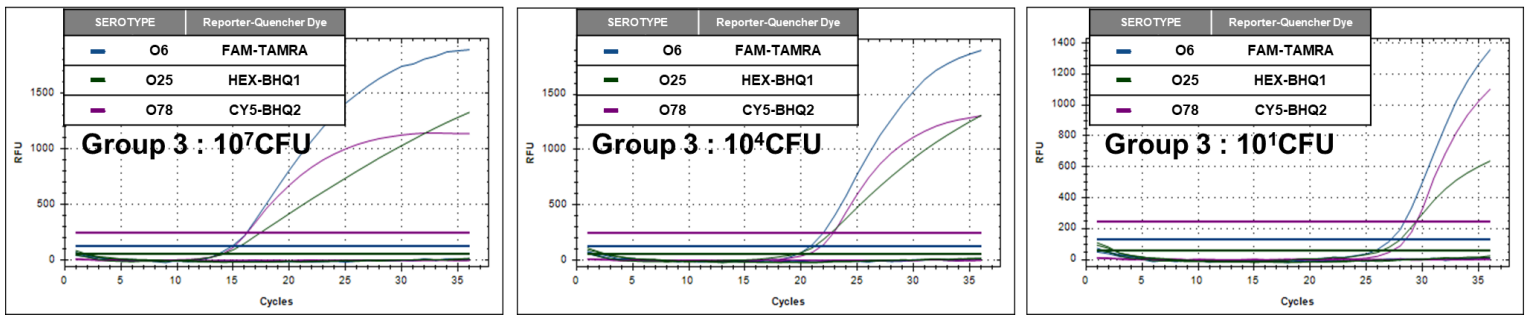


Fig. 11. Food application test of the target *E. coli* strain at three different CFU numbers (10^7, 10^4, and 10^1 CFUs per sample) for O-serotyping in food samples.

(A) Group 1: serotyping of O145, O157, and O26. (B) Group 2: serotyping of O103 and O111.

C



D

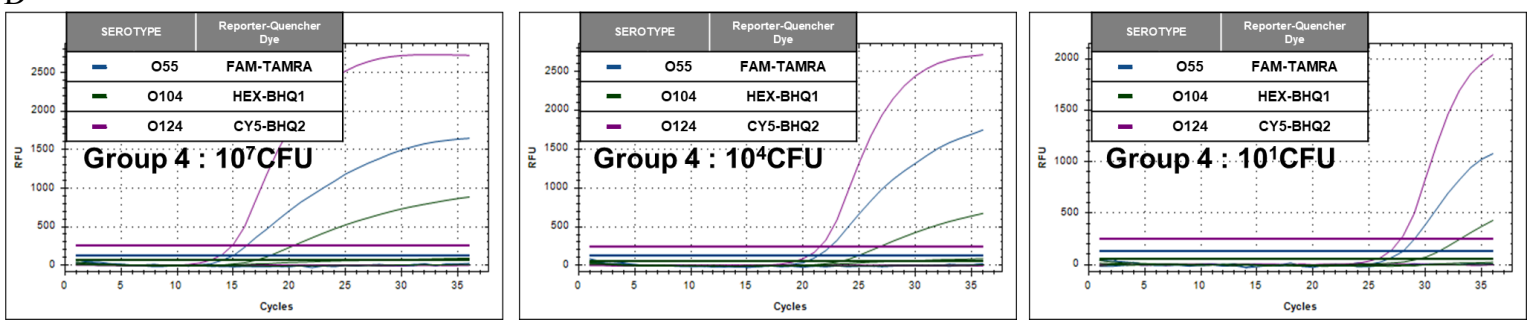


Fig. 11. **Cont.** Food application test of the target *E. coli* strain at three different CFU numbers (10^7, 10^4, and 10^1 CFUs per sample) for O-serotyping in food samples.

(C) Group 3: serotyping of O6 and O25. (D) Group 4: serotyping of O55 and O104.

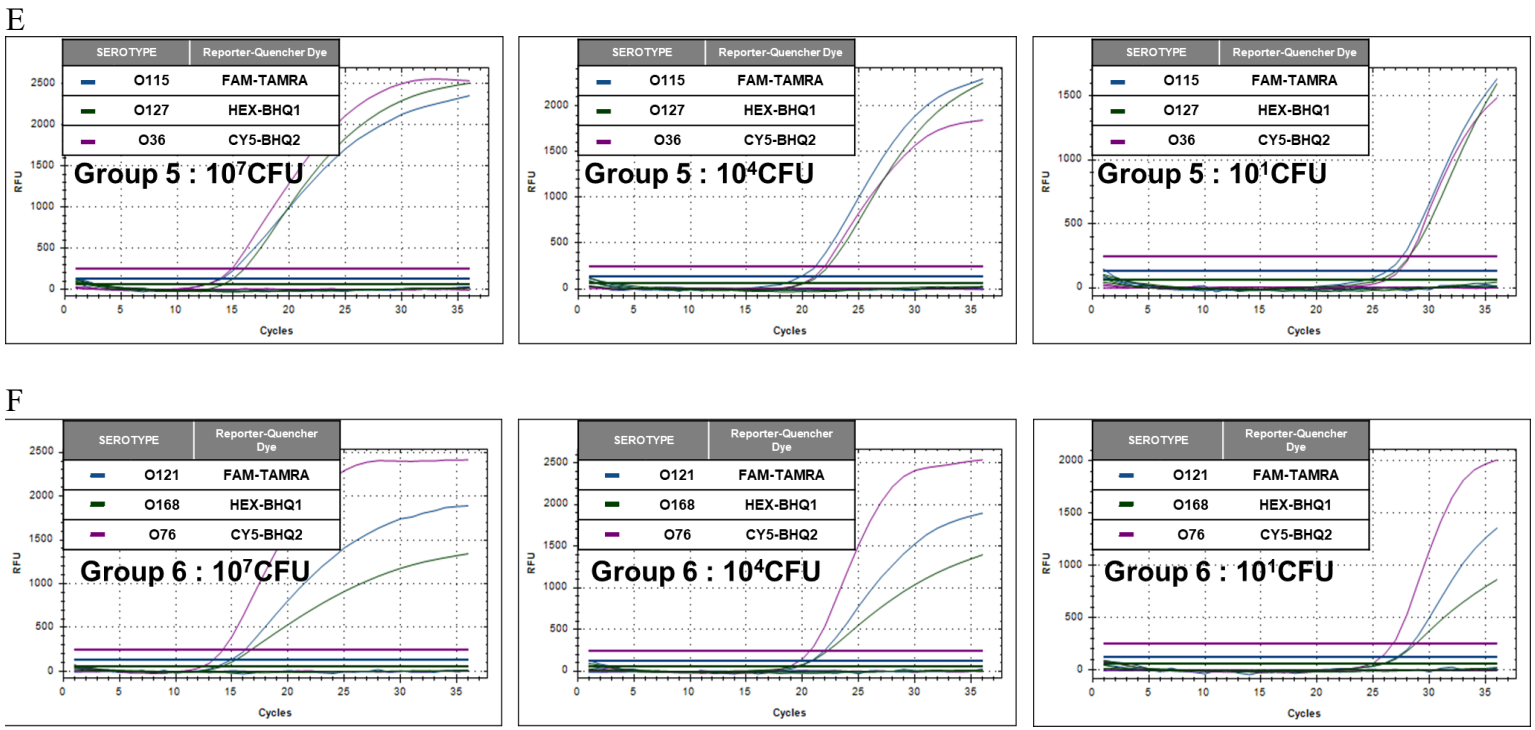
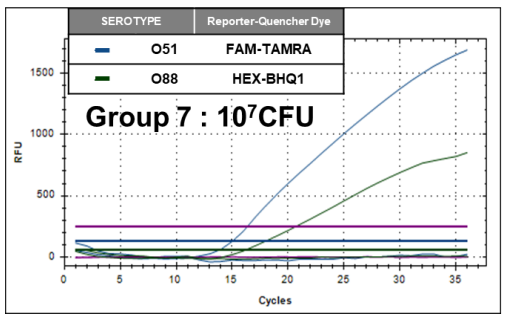
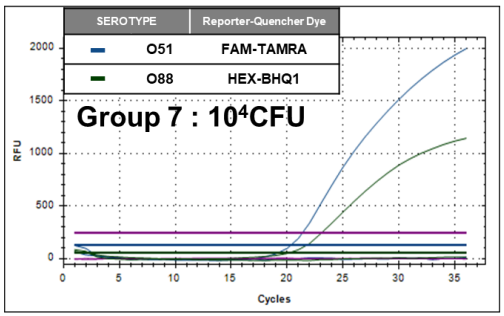


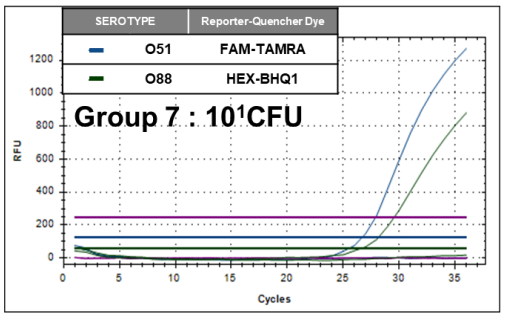
Fig. 11. **Cont.** Food application test of the target *E. coli* strain at three different CFU numbers (10^7, 10^4, and 10^1 CFUs per sample) for O-serotyping in food samples.

(E) Group 5: serotyping of O115, O127, and O36. (F) Group 6: serotyping of O121, O168, and O76.

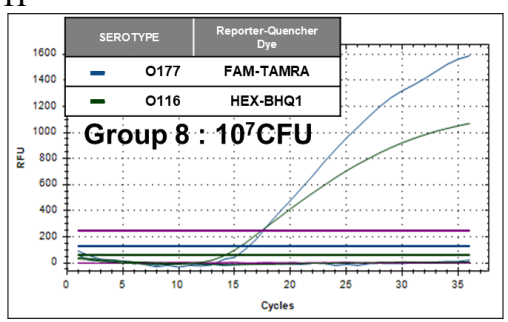
G

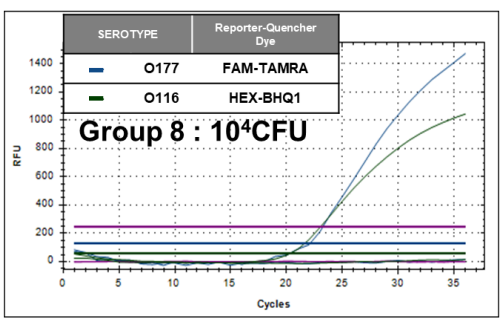






H





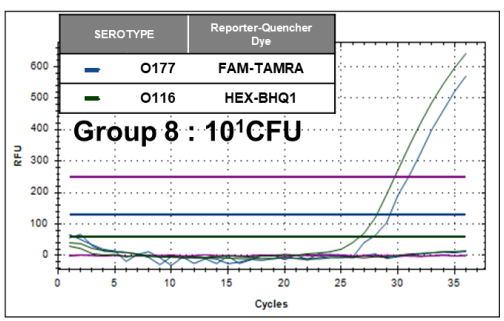


Fig. 11. **Cont.** Food application test of the target *E. coli* strain at three different CFU numbers (10^7, 10^4, and 10^1 CFUs per sample) for O-serotyping in food samples.

(G) Group 7: serotyping of O51 and O88. (H) Group 8: serotyping of O177 and O116.

4. Discussion

Foodborne outbreaks caused by pathogenic *E. coli* occur worldwide every year, affecting numerous patients and causing economic losses. Previous studies have shown that *E. coli* pathogenicity and serotypes are highly correlated (Cho et al., 2010; Tamura et al., 1996). Therefore, to prevent severe foodborne outbreaks due to the consumption of foods contaminated with specific O-serotype pathogenic *E. coli*, the rapid and accurate identification of *E. coli* serotypes is essential. The conventional serotyping method, antiserum serotyping, is expensive, time-consuming, and not always definitive, and the *E. coli* genome sequence database is still insufficient for *in silico* serotyping, which is also expensive and time-consuming when a WGS is obtained via NGS, even though *in silico* serotyping is quick and accurate. In this study, a novel multiplex real-time PCR serotyping method was developed and optimized for the identification of 21 different *E. coli* O-serotypes. An evaluation test using 54 *E. coli* strains comprising 11 *E. coli* O-serotype reference strains and 43 environmental isolates provided results that were perfectly matched with those of antiserum serotyping and *in silico* serotyping, indicating that the simple and inexpensive multiplex real-

time PCR serotyping method reported here could replace traditional serotyping methods without increasing false-positive reactions. Interestingly, the PCR reaction time for O-serotyping was <40 min and the detection limit was 1–10 pg DNA, suggesting that identification via the developed method is rapid and highly sensitivity. Moreover, the novel method requires only 8 reactions for the multiple detection of 21 different *E. coli* O-serotypes. Therefore, this new method can be considered the optimal O-serotyping tool for the determination of *E. coli* O-serotypes produced to date. Consequently, the method could be applied for the prevention of foodborne outbreaks caused by the contamination of food by pathogenic *E. coli* with specific O-serotypes.

To date, 186 *E. coli* O-serotypes have been reported (Fratamico, DebRoy, Liu, Needleman, Baranzoni, & Feng, 2016; Sun et al., 2011). However, the multiplex real-time PCR serotyping method developed in this study covers only 21 O-serotypes in 8 groups; thus, the development and optimization of additional primer–probe sets is still required. Even if every primer–probe set required to cover the 186 O-serotypes was developed and optimized, approximately 60–70 groups would be needed, suggesting that 60–70 real-time PCR reactions for O-

serotyping should be performed to differentiate and determine the 186 known O-serotypes.

One alternative to multiplex real-time PCR, the NGS panel analysis method, was initially developed for clinical cancer diagnosis and the determination of genetically modified organisms via the meta-shotgun NGS of various target genes (Arulandhu et al., 2018; López-Reig et al., 2019). The NGS panel analysis method benefits from the multiple detection and identification of target genes in one reaction. NGS panel analysis has been used on a few occasions for the multiple detection and determination of foodborne pathogens in various foods and environmental samples (Bridier, 2019; Mira Miralles, Maestre-Carballa, Lluesma-Gomez, & Martinez-Garcia, 2019). Although the NGS panel analysis method has not been fully optimized, if it could be applied for the identification and determination of the 186 known *E. coli* O-serotypes in one reaction, it would serve as a useful tool for limiting foodborne outbreaks. However, NGS panel analysis has limitations such a lower detection limit and lower sensitivity than multiplex real-Time PCR. To overcome these limitations, NGS panel primer sets could be optimized for sensitivity, and a real-time NGS sequencing technology, such as a nanopore sequencer, could be used for rapid detection and

identification. However, until NGS panel analysis becomes available, multiplex real-time PCR should be considered the optimum method for detecting and identifying *E. coli* O-serotypes.

5. References

Al-Soud, W. A. & Rådström, P. (2000). Effects of amplification facilitators on diagnostic PCR in the presence of blood, feces, and meat. *Journal of Clinical Microbiology, 38*(12), 4463-4470.

Arulandhu, A. J., van Dijk, J., Staats, M., Hagelaar, R., Voorhuijzen, M., Molenaar, B., van Hoof, R., Li, R., Yang, L., Shi, J., Scholtens, I., Kok, E. (2018). NGS-based amplicon sequencing approach; towards a new era in GMO screening and detection. *Food Control, 93*, 201-210.

Bazaka, K., Crawford, R. J., Nazarenko, E. L., & Ivanova, E. P. (2011). Bacterial extracellular polysaccharides. In D. Linke & A. Goldman (Eds.), *Bacterial Adhesion* (pp. 213-226). Springer New York, NY.

Bridier, A. (2019). Exploring foodborne pathogen ecology and antimicrobial resistance in the light of shotgun metagenomics. In A. Bridier (Ed.), (pp. 229-245). Humana New York, NY.

Burger, R. (2012). EHEC O104: H4 in Germany 2011: Large outbreak of bloody diarrhea and haemolytic uraemic syndrome by Shiga toxin-producing *E. coli* via contaminated food.

Chen, Y. H. & Hsueh, P. R. (2011). European *Escherichia coli* O104: H4 outbreak. *Journal of the Formosan Medical Association, 110*(10), 611-612.

Cho, S. H., Oh, K. H., Kim, S. H., Oh, H. B., & Park, M. S. (2010). Distribution of virulence genes and their association of serotypes in pathogenic *Escherichia coli* isolates from diarrheal patients in Korea. *Osong Public Health and Research Perspectives, 1*(1), 29-35.

DebRoy, C., Fratamico, P. M., Roberts, E., Davis, M. A., & Liu, Y. (2005). Development of PCR assays targeting genes in O-antigen gene clusters for detection and identification of *Escherichia coli* O45 and O55 serogroups. *Applied and Environmental Microbiology, 71*(8), 4919-4924.

DebRoy, C., Fratamico, P. M., Yan, X., Baranzoni, G., Liu, Y., Needleman, D. S., . . . Swimley, M. (2016). Comparison of O-antigen gene clusters of all O-serogroups of Escherichia coli and proposal for adopting a new nomenclature for O-typing. PLoS One, 11(1), e0147434.

Dewey-Mattia, D., Manikonda, K., Hall, A. J., Wise, M. E., & Crowe, S. J. (2018). Surveillance for foodborne disease outbreaks—United

States, 2009–2015. *Morbidity and Mortality Weekly Report. Surveillance Summaries, 67*(10), 1-11.

Fratamico, P. M., DebRoy, C., Liu, Y., Needleman, D. S., Baranzoni, G. M., & Feng, P. (2016). Advances in molecular serotyping and subtyping of *Escherichia coli*. *Frontiers in Microbiology, 7*, 644.

Iguchi, A., Iyoda, S., Seto, K., Morita-Ishihara, T., Scheutz, F., Ohnishi, M., & Pathogenic *E. coli* Working Group in Japan (2015). *Escherichia coli* O-genotyping PCR: A comprehensive and practical platform for molecular O serogrouping. *Journal of Clinical Microbiology, 53*(8), 2427-2432.

Hoon-Jae Jeong. (2020). Determiantion of *Escherichia coli* O-serotype Using Real-time PCR. Master's thesis, Kyung Hee University, Seoul.

Joensen, K. G., Tetzschner, A. M., Iguchi, A., Aarestrup, F. M., & Scheutz, F. (2015). Rapid and easy *in silico* serotyping of *Escherichia coli* isolates by use of whole-genome sequencing data. *Journal of Clinical Microbiology, 53*(8), 2410-2426.

Kaper, J. B., Nataro, J. P., & Mobley, H. L. (2004). Pathogenic *Escherichia coli*. *Nature Reviews Microbiology, 2*(2), 123-140.

Kauffmann, F. (1947). The serology of the coli group. *Journal of*

Immunology, 57(1), 71-100.

Lane, D. (1991). Nucleic acid sequencing techniques in bacterial systematics.

Lee, S. W., Lee, B. K., Lee, Y. J., Lee, H. S., Jung, S. C., Sun, K. H., & Choi, B. Y. (2005). A study on epidemiological characteristics and control methods of EHEC infection in Korea. *Korean Journal of Epidemiology, 27*(1), 37-52.

López-Reig, R., Fernández-Serra, A., Romero, I., Zorrero, C., Illueca, C., García-Casado, Z., Poveda, A., López-Guerrero, J. A. (2019). Prognostic classification of endometrial cancer using a molecular approach based on a twelve-gene NGS panel. *Scientific Reports, 9*(1), 18093.

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.

Ørskov, F. & Ørskov, I. (1984). 2 Serotyping of *Escherichia coli*. In T. Bergan (Ed.), *Methods in Microbiology* (pp. 43-112). Academic Press.

Page, A. J., Cummins, C. A., Hunt, M., Wong, V. K., Reuter, S., Holden, M. T., . . . Parkhill, J. (2015). Roary: rapid large-scale prokaryote pan genome analysis. Bioinformatics, 31(22), 3691-3693.

Rangel, J. M., Sparling, P. H., Crowe, C., Griffin, P. M., & Swerdlow, D. L. (2005). Epidemiology of *Escherichia coli* O157: H7 outbreaks, United States, 1982–2002. *Emerging Infectious Diseases, 11*(4), 603-609.

Reeves, P. (1995). Role of O-antigen variation in the immune response. *Trends in Microbiology, 3*(10), 381-386.

Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. Bioinformatics, 30(14), 2068-2069.

Smith, C., Griffiths, A., Allison, S., Hoyano, D., & Hoang, L. (2022). *Escherichia coli* O103 outbreak associated with minced celery among hospitalized individuals in Victoria, British Columbia, 2021. *Canada Communicable Disease Report = Releve des Maladies Transmissibles au Canada, 48*(1), 46-50.

Sun, Y., Wang, M., Liu, H., Wang, J., He, X., Zeng, J., Guo, X., Li, K., Cao, B., Wang, L. (2011). Development of an O-antigen serotyping scheme for *Cronobacter sakazakii*. *Applied and Environmental Microbiology, 77*(7), 2209-2214.

Tamura, K., Sakazaki, R., Murase, M., & Kosako, Y. (1996). Serotyping and categorisation of *Escherichia coli* strains isolated between 1958 and 1992 from diarrhoeal diseases in Asia. *Journal of Medical Microbiology, 45*(5), 353-358.

Wang, L. & Reeves, P. R. (1998). Organization of *Escherichia coli* O157 O antigen gene cluster and identification of its specific genes. *Infection and Immunity, 66*(8), 3545-3551.

Wang, X. & Quinn, P. J. (2010). *Endotoxins: Structure, function and recognition* (vol. 53). Springer Science & Business Media.

West, N. P., Sansonetti, P., Mounier, J., Exley, R. M., Parsot, C., Guadagnini, S., Prévost, M. C., Prochnicka-Chalufour, A., Delepierre, M., Tanguy, M., Tang, C. M. (2005). Optimization of virulence functions through glucosylation of Shigella LPS. *Science, 307*(5713), 1313-1317.

국문초록

병원성 대장균에 의해 발생하는 식품 매개 감염병은 전 세계적으로 일어나고 있으며, 경제적 손실과 관련 질병을 유발하고, 심지어 사망을 야기하기도 한다. 이전의 연구들은 대장균의 병원성과 혈청형이 높은 상관관계가 있다는 것을 보여주었기 때문에, 대장균 혈청형의 신속하고 정확한 식별은 심각한 식품 매개 감염병 발병을 예방하는 데 필수적이다. 본 연구에서는 real-time PCR 을 이용한 새로운 O-혈청형 동정 방법을 개발하여, 기존의 O-혈청형 동정 방법인 antiserum 방법과 *in silico* 동정 방법의 한계를 극복하였다. 즉, 높은 감도와 특이성을 가지며 낮은 검출 한계를 극복하고 감도의 손실 없이 O-혈청형 검출 및 동정 시간을 단축하였다. 구체적으로, 두 개의 O-혈청형 특이 유전자인 *wzy* 와 *wzx* 가 서로 다른 O-혈청형을 결정하기 위한 21 개의 프라이머/프로브 세트를 설계하는데 사용되었다. 그런 다음 11 주의 reference 대장균과 43 주의 환경 분리 대장균을 사용해 방법을 평가하였다. O-혈청형 동정 방법은 1-10 pg

DNA 의 낮은 검출 한계를 가지며 40 분 안으로 완료되었다. 또한 PCR 혈청형 동정 방법은 기존 혈청형 동정 방법과 동일한 결과를 보여 효과적임을 확인할 수 있었다. 식품 적용 테스트 또한 완료되었다. 이 방법은 다진 소고기 샘플에서 O-혈청형을 샘플 당 10^1 CFU 에서도 식별할 수 있었다. 본 연구의 결과에 따르면, 새로운 이 방법은 현재까지 대장균 O-혈청형을 동정하기 위한 최적의 방법으로 간주될 수 있다. 개발된 real-time PCR 혈청형 동정 방법은 빠르고, 저렴하고, 정확하게 대장균 O-혈청형을 식별할 수 있으므로, 식중독 발생을 예방하는 데 도움이 될 수 있다.

주요어: **대장균, O-혈청형, 혈청형 동정, 다중 실시간 중합효소연쇄반응, 생물정보학**

학번: 2021-29876