Rapid Real-Time Polymerase Chain Reaction for Salmonella Serotyping Based on Novel Unique Gene Markers by Pangenome Analysis

Seung-Min Yang, Eiseul Kim, Dayoung Kim, Hyeon-Be Kim, Jiwon Baek, Seyoung Ko, Donghyuk Kim, Hyunjin Yoon, and Hae-Yeong Kim

An accurate diagnostic method for Salmonella serovars is fundamental to preventing the spread of associated diseases. A diagnostic polymerase chain reaction (PCR)-based method has proven to be an effective tool for detecting pathogenic bacteria. However, the gene markers currently used in real-time PCR to detect Salmonella serovars have low specificity and are developed for only a few serovars. Therefore, in this study, we explored the novel unique gene markers for 60 serovars that share similar antigenic formulas and show high prevalence using pangenome analysis and developed a real-time PCR to detect them. Before exploring gene markers, the 535 Salmonella genomes were evaluated, and some genomes had serovars different from the designated serovar information. Based on these analyses, serovar-specific gene markers were explored. These markers were identified as genes present in all strains of target serovar genomes but absent in strains of other serovar genomes. Serovar-specific primer pairs were designed from the gene markers, and a real-time PCR method that can distinguish between 60 of the most common Salmonella serovars in a single 96-well plate assay was developed. As a result, real-time PCR showed 100% specificity for 199 Salmonella and 29 non-Salmonella strains. Subsequently, the method developed was applied successfully to both strains with identified serovars and an unknown strain, demonstrating that real-time PCR can accurately detect serovars of strains compared with traditional serotyping methods, such as antisera agglutination. Therefore, our method enables rapid and economical Salmonella serotyping compared with the traditional serotyping method.

Keywords: Salmonella, serotyping, pangenome analysis, detection, real-time PCR, gene marker

INTRODUCTION

The genus Salmonella, the causative agent of foodborne salmonellosis, can infect both animals and humans, leading to public health problems and economic loss (Gand M. et al., 2020a). Most Salmonella infections are caused by consuming contaminated water or food (Kasturi, 2020). Currently, Salmonella is divided into two species and six subspecies. Serotypes are further classified...
into more than 2,600 serovars following the White–Kauffman–Le Minor scheme, using antigenic agglutination reactions to three cell-surface antigens of somatic O, and flagellar H antigens denoted as H1 and H2 (Grimont and Weill, 2007; Yachison et al., 2017; Zhang et al., 2019; Gand M. et al., 2020a). As a reliable surveillance protocol is critical for detecting outbreaks or preventing their spread, using a differential serotyping method that identifies serogroups and serovars of Salmonella isolates from causative agents is important (Kasturi, 2020).

Traditional serotyping methods require numerous antisera, are labor-intensive, time-consuming, and complicated, and may produce ambiguous results (Hong et al., 2008; Xiong et al., 2017). Some isolates do not express antigens because of a single-nucleotide mutation in the genome, and some require multiple passes through semisolid media to enhance the flagella antigen expression (Ibrahim and Morin, 2018). As a result of these limitations, serotyping by antigenic agglutination and biochemical tests has been replaced by molecular serotyping (Zhang et al., 2019). For the rapid diagnosis or tracking of Salmonella serovars, epidemiological investigations have been conducted by molecular serotyping analysis based on pulsed-field gel electrophoresis, plasmid profiles, and polymerase chain reactions (PCRs) (Ozdemir and Acar, 2014; Gad et al., 2018; Yang et al., 2021). Notably, the PCR-based method is widely used for early diagnosis because it can diagnose a few bacteria in the specimen and can yield rapid results. However, as the PCR-based method for serotyping mainly uses markers within genes responsible for somatic and flagellar antigen expression, these genes do not detect strains that share the same antigenic formula. Thus, although PCR is reasonably rapid and inexpensive compared with conventional serotyping methods, the limiting factors of this assay are that molecular serotyping does not diagnose numerous serovars and focuses mainly on the most common serovars, such as Salmonella Typhimurium and Enteritidis (Ibrahim and Morin, 2018).

Previous studies have identified gene markers, such as fimA, hilA, invA, and ttr, that are useful in detecting Salmonella species (Laing et al., 2017; Gand M. et al., 2020b; Kreitlow et al., 2021). Moreover, for serovar identification, gene markers based on allelic variations in O and H antigen genes were used (Laing et al., 2017; Gand M. et al., 2020b). However, this gene marker cannot distinguish isolates that share the same antigenic formula. To overcome this problem, some studies have identified markers, such as STM0292, STM4200, STM4493, and STM2235 specific to Typhimurium and SEN1392 specific to Enteritidis, using comparative genomics (Akiba et al., 2011; Liu et al., 2012). These gene markers have specificity but have been screened using a limited number of genomes and cannot detect various serovars in a single reaction (Heymans et al., 2018; Ibrahim and Morin, 2018).

Advances in whole-genome sequencing technology have improved the understanding of the species and subtypes of pathogenic bacteria, and can provide information on the virulence of the underlying pathogenesis. Data obtained using whole-genome sequencing technology can be used to confirm paths of disease transmission and provide information on potential outbreaks (Ibrahim and Morin, 2018; Diep et al., 2019; Cooper et al., 2020). Therefore, whole-genome sequencing is currently used as a technique to obtain reliable and rapid serovar information. Recently, multiple in silico tools have been developed to determine Salmonella serovars from whole-genome sequence data (Zhang et al., 2015, 2019; Yoshida et al., 2016). SeqSero and Salmonella In Silico Typing Resource (SISTR), which can infer serovar predictions from analyses of somatic O and flagellar H antigens derived from whole-genome sequence data, are representative in silico analysis tools (Uelze et al., 2020). Whole-genome sequencing and in silico tools have many advantages in serotyping, for example, they reveal detailed genetic information on the characteristics of isolates and accurate serovar predictions, provided the database is sufficient (Ibrahim and Morin, 2018). However, whole-genome sequence-based method must be considered against costs associated with genome sequencing to analyze the many samples. This method also requires trained researchers with a specific skill set (Mellmann et al., 2017). To overcome the limitations of genome sequencing, primer sets for PCR targeting serovar-specific genes were recently developed. They show accuracy for serovar detection but still cannot detect a large number of serovars (Shang et al., 2021; Ye et al., 2021).

The purpose of this study is to evaluate Salmonella genomes by in silico serotyping, to select novel serovar-specific gene markers based on pan genome analysis, and to develop a real-time PCR method that can distinguish between 60 of the most common Salmonella serovars in a single 96-well plate by detecting unique serovar-specific gene markers.

MATERIALS AND METHODS

In silico Serotyping

This study selected 60 serovars, which have been frequently isolated worldwide, are essential for public health, and are difficult to diagnose using traditional serotyping methods. The target serovars are as follows: Aberdeen, Agona, Albany, Anatum, Bareilly, Berta, Blockley, Braenderup, Brandenburg, Cerro, Choleraesuis, Corvallis, Derby, Dublin, Elisabethville, Enteritidis, Gallinarum, Give, Hadar, Heidelberg, I 4,[5],12:i:-, Infantis, Javiana, Kedougou, Kentucky, Kottbus, Litchfield, Livingstone, London, Manhattan, Mbandaka, Meleagrisidis, Menston, Minnesota, Mississippi, Montevideo, Muenchen, Muenster, Newington, Newport, Ohio, Oranienburg, Panama, Paratyphi B, Poona, Reading, Rissen, Saintpaul, Schwarzengrund, Senftenberg, Singapore, Stanley, Tennessee, Thompson, Typhi, Typhimurium, Uganda, Vinohrady, Virchow, and Weltevreden. The 535 Salmonella assembled genome sequences representing 60 serovars were obtained from the National Center for Biotechnology Information (NCBI) (Supplementary Table 1).

All genomes used in this study were subjected to in silico serotyping using SeqSero2 version 1.2.1 and SISTR version 1.1.1. SeqSero used the assembled genome sequence as the input and the sequence was analyzed using Python code (Zhang et al., 2015, 2019). Also, serotyping and core-genome multilocus sequence typing (cgMLST) of whole-genome sequences were analyzed using a SISTR command-line tool.
Pangenome Analysis and Discovery of Unique Gene Markers

The pangenome was analyzed using workflow for microbial pangenomic analysis with the Anvi’o package version 6.1 (Eren et al., 2015). Briefly, the assembled genome sequences were used as input and clustered based on the similarity of the amino acid sequences by the Markov cluster algorithm and NCBI’s blastp algorithm according to the developer recommendations (Kayansamruaj et al., 2019). Then the pangenome result was visualized using anvi-display-pan code of Anvi’o, and the genomes were organized based on the pan gene cluster frequencies.

The unique gene of each serovar was obtained using a Bacterial Pan Genome Analysis (BPGA) version 1.3 (Chaudhari et al., 2016). The annotated protein in fasta-format file was used as input, and the pangenome was analyzed with the default value (cut-off: 50%). All genomes were compiled into separate local databases, which include the core-genome composed of proteins common to genomes of the target serovar, and the pangenome is composed of entire proteins of all genomes. The unique genes of each serovar were gathered by comparing the pan and core-genome databases. The extracted unique genes were aligned with 65,815,883 sequences using the Basic Local Alignment Search Tool (BLAST). To verify the unique gene markers, the unique gene of each serovar was aligned with 535 Salmonella genomes using USEARCH version 9.0 (Edgar, 2010). Afterward, serovar-specific primer pairs were designed from selected gene markers. To increase the efficiency of primer pairs, the length was 18–30 base pairs, the guanine–cytosine (GC) content was designed according to the manufacturer’s instructions. The concentration of each primer pair was run independently in a single 96-well plate (Supplementary Table 2). The genomes used in the in silico PCR come from NCBI and EnteroBase.

Cultured Bacterial Strains and DNA Extraction

The 199 Salmonella strains, 33 Salmonella species or Salmonella enterica strains, and 29 non-Salmonella strains used in this study are presented in Supplementary Table 3. All bacterial strains were grown in TSB at 37°C for 18 h under aerobic conditions. The cultured strains were collected by centrifugation at 13,600 × g for 5 min. Then, genomic DNA was extracted from the pellet using the DNeasy Blood & Tissue Kit (QIAGEN, Hilden, Germany), according to the manufacturer’s instructions. The concentration and purity of the genomic DNA extracted were measured using the MaestroNano® spectrophotometer (MaestroLog, Las Vegas, NV, United States).

Specificity and Accuracy for Developed Primer Pairs

Each 20-µl real-time PCR reaction mixture contained 500 nM of each primer pair, 10 µl of 2× Thunderbird SYBR® qPCR Mix (Toyobo, Osaka, Japan), and 20 ng of template DNA. To avoid false-negative results, an internal standard targeting the bacterial 16S rRNA gene fragment was used (Aboutalebian et al., 2021). Amplification was conducted in a 7500 Real-Time PCR System (Applied Biosystems, Foster City, CA, United States), with an initial denaturation for 2 min at 95°C, followed by 30 cycles of 95°C for 5 s and 60°C for 30 s. The melting curve was generated according to the following conditions: 95°C for 15 s, 60°C for 1 min, 95°C for 30 s, and 60°C for 15 s. The specificity of the primer pairs developed was tested against each Salmonella strain as well as each non-Salmonella reference strain. Primer pairs cross reactivity across all of the serovars were evaluated. To evaluate the accuracy, genomic DNA for each serovar were serially diluted, and then standard curves were generated in triplicate using diluted DNA from 0.0002 to 20 ng. The results obtained were analyzed using 7500 software version 2.3 (Applied Biosystems).

Evaluation of Real-Time Polymerase Chain Reaction With a Single 96-Well Plate

A real-time PCR method was developed that can distinguish between 60 of the most common Salmonella serovars in a single 96-well plate. The real-time PCR was designed so that each primer pair was run independently in a single 96-well plate (Supplementary Figure 1). Each well contained different primer pair, and the genomic DNA from a single isolate was added to each well. The serovar of strain was determined as serovar corresponding to the primer pair included in the well in which amplification occurred. The real-time PCR developed in this study was evaluated using 189 Salmonella strains whose serovars were confirmed through antisera agglutination and 33 strains whose serovars were unknown. Each strain was tested against each of the primer pairs. For serovar diagnosis of isolates, genomic DNA of the isolate was added to each well of the reaction plate containing different primer pair and 2× Thunderbird SYBR® qPCR Mix (Toyobo). Then, real-time PCR was conducted in the 7500 Real-Time PCR System (Applied Biosystems). The conditions were the same as those described in the previous “Specificity and Accuracy for Developed Primer Pairs” section.

Traditional Serotyping Using Antisera Agglutination

The antigenic formulas of strains were determined by the World Health Organization Collaborating Center for Reference and Research on Salmonella, located at the Pasteur Institute, and the serovar name was assigned by the White–Kauffmann–Le Minor scheme (Grimont and Weill, 2007). All strains were cultured in brain heart infusion (BD Difco, Sparks, MD, United States) and motility GI medium (BD Difco) to determine somatic and flagellar antigens. The somatic antigen was confirmed by the slide agglutination reaction using antisera (BD Difco). The flagellar phase was activated by the motility test and fixed by treatment with 0.6% formalin. The flagellar antigen was determined by an aggregation reaction in glass test tubes by mixing with an antisera solution (BD Difco). The serotyping

1http://in-silico.ehu.es/PCR/
results of antisera agglutination were compared with the results of the real-time PCR method.

RESULTS AND DISCUSSION

**In silico Serotyping**

Whole-genome sequencing data have been used for serotype diagnosis or subtyping of *Salmonella* (Zhang et al., 2019; Uelze et al., 2020). SeqSero is a web-based serotyping tool that can predict many *Salmonella* serovars using whole-genome sequence data based on a database of *Salmonella* serovar determinants (Ibrahim and Morin, 2018). This tool extracts the relevant genomic regions of cell-surface antigens, such as the rfb gene cluster to determine the O antigen and the fliC and fliB genes to determine H1 and H2 antigens, from the genome assemblies or raw sequencing reads and aligns these genes to the curated database using BLAST (Zhang et al., 2015; Ibrahim and Morin, 2018). This software then determines the serovar according to the White–Kauffmann–Le Minor scheme.

In this study, 511 genomes (95.51%) were predicted as correct serovars by SeqSero, and the remaining 24 genomes (4.49%) were predicted as incorrect serovars (Supplementary Table 4). Of the 24 genomes that produced incorrect serovars, eight genomes were predicted to be serotypes inconsistent with serovar nomenclature reported to NCBI, and 13 genomes indicated two or more serotypes (Table 1). This shows that as the serotype by SeqSero is determined only by the O and H antigens, these genomes show that two or more serovars were indicated for similar serovars, such as Gallinarum or Enteritidis, Paratyphi C or Choleraesuis or Typhisuis, and Albany or Duesseldorf. Additionally, some genomes (n = 4) were determined to be partial serovars lacking O or H1 or H2-antigens, and accurate serovar information could not be extracted. The reason for this appears to be the failure to extract the serovar determinant from the assembled genomes (Zhang et al., 2015).

Salmonella In Silico Typing Resource also determines the serovar according to the White–Kauffmann–Le Minor scheme, based on their databases of *Salmonella* serotype determinants (wzx/wzy, fliC, and fliB alleles) (Yoshida et al., 2016). To resolve the ambiguous serovar designations resulting from antigen determination, SISTR uses the novel 330 locus cgMLST analysis, and together, these two determinants are used to provide an overall serovar prediction (Yoshida et al., 2016). As a result of SISTR analysis, 526 genomes (98.32%) were predicted as correct serovars (Supplementary Table 5), whereas nine genomes (1.68%) were predicted as incorrect serovars (Table 1).

In *silico* serotyping analyses of the genomes obtained from the NCBI showed 95.51 and 98.32% accuracy on SeqSero and SISTR platforms, respectively. These data are consistent with the success rates of each platform reported in previous studies (Zhang et al., 2015; Yoshida et al., 2016). However, some limitations of whole-genome sequencing also exist. Previous studies have reported that SeqSero provided two possible serovars that share the same antigenic formula but differed in the minor O antigen factor (Ibrahim and Morin, 2018). In this study, serovars such as Anatum (3,10:e,h:1,6) and Newington (Anatum var. 157, 3,10,15:e,h:1,6) shared similar antigenic formulas, yielding only Anatum serovars. Moreover, *in silico* serotyping tools could not predict some genomes. These unpredicted serovars may infrequently be separated, or possibly there were some gaps in their databases (Ibrahim and Morin, 2018).

**Genome Evaluation**

Previous studies have reported that the NCBI has often misclassified genomes, so the genome evaluation should be conducted before using the genome obtained from the NCBI (Kim et al., 2020, 2021). Therefore, the genomes used in this study were evaluated to prevent incorrect results obtained in selecting unique serovar genes.

| Serovars | Predicted serovar by SeqSero2 | Predicted serovar by SISTR |
|----------|-------------------------------|---------------------------|
| Albany ATCC 51960 | Albany or Duesseldorf (8:z4,z24:-) | Albany |
| Albany sg_wt5 | Albany or Duesseldorf (8:z4,z24:-) | Albany |
| Albany CFSAN103854 | Albany or Duesseldorf (8:z4,z24:-) | Albany |
| Albany R17.5974 | Albany or Duesseldorf (8:z4,z24:-) | Albany |
| Albany R16.0556 | Albany or Duesseldorf (8:z4,z24:-) | Albany |
| Choleraesuis SC-B67 | Paratyphi C or Choleraesuis or Typhisuis (7:C:1,5) | Choleraesuis |
| Choleraesuis A50 | Paratyphi C or Choleraesuis or Typhisuis (7:C:1,5) | Choleraesuis |
| Choleraesuis ATCC 10708 | Paratyphi C or Choleraesuis or Typhisuis (7:C:1,5) | Choleraesuis |
| Choleraesuis C500 | Paratyphi C or Choleraesuis or Typhisuis (7:C:1,5) | Choleraesuis |
| Derby 2013LSAL02374 | Typhimurium (4:i:1,2) | Typhimurium |
| Enteritidis 92-0392 | Gallinarum or Enteritidis (9:g,m:-) | Gallinarum |
| Gallinarum ATCC 9120 | Gallinarum or Enteritidis (9:g,m:-) | Gallinarum |
| Gallinarum 13026 | Gallinarum or Enteritidis (9:g,m:-) | Gallinarum |
| Gallinarum 19945 | Gallinarum or Enteritidis (9:g,m:-) | Gallinarum |
| Gallinarum S4037-07 | Gallinarum or Enteritidis (9:g,m:-) | Gallinarum |
| Give SS-487 | Give (3,10:l,v:1,7) | E1:l,v:z13,z28 |
| London CFSAN001081 | Typhimurium (4:i:1,2) | Typhimurium |
| Mbandaika CFSAN076213 | Typhimurium (4:i:1,2) | Mbandaika |
| Mississippi SAL-19-VL-SD-NC-0011 | I :b:1,5 (-:b:1,5) | Mississippi |
| Newington 261388 | Anatum (3,10:e,h:1,6) | Anatum |
| Newington 95006 | Anatum (3,10:e,h:1,6) | Anatum |
| Paratyphi B SAR061 | Agona (4:f,g:z4,-) | Agona |
| Typhimurium TW-Strm6 | Enteritidis (9:g,m:-) | Enteritidis |
| Typhimurium FORC50 | Enteritidis (9:g,m:-) | Enteritidis |
| Typhimurium FORC98 | Tennessee (7:z29:-) | Tennessee |
Yang et al. Real-Time PCR for Salmonella Serotyping

As a result of phylogeny based on the pangene cluster frequencies among the 535 genomes, most genomes were clustered according to serovar types (Figure 1). Serovars with similar antigenic formula types, such as Typhimurium and I 4,[5],12:i:-, were adjacent but clustered into different groups. These two serovars differ by only one flagellar antigen (1,4,[5],12:i:1,2 vs. 4,[5],12:i:-) in their antigenic formulas. However, some Enteritidis, Typhimurium, and Paratyphi B strains were clustered with different serovars. Paratyphi B SARA61 was clustered into Agona; Typhimurium TW-Stm6, FORC50, and FORC098 were clustered into I 4,[5],12:i:-, Enteritidis, and Tennessee, respectively; Enteritidis 92-0392 was clustered into Typhimurium. Most genomes were clustered with the same serovar group because of in silico serotyping analyses. This is consistent with a previous study that reported that similar or rare serotypes could produce unpredicted serovar results using the SeqSero database (Ibrahim and Morin, 2018).

Identification of Serovar-Specific Gene Markers
A total of 2,440,535 genes yielded a pangenome size of 15,853 genes. The core genome comprises 1,215 genes; the accessory genome, 11,156 genes; and the unique genome, 3,482 genes. The core genomes for each serovar included from 3,065 to 4,702 genes, and unique genomes contained from 1 to 160 genes. The unique genes considered specific for each serovar were identified by BLAST analysis, and genes rarely present in other bacteria and specific to the serovar were selected. Further, a unique gene marker of each serovar was finally selected considering the GC
### Table 2 | Information of gene markers obtained pangenome analysis.

| Serovar          | Protein name                        | Accession no. | Size (bp) |
|------------------|-------------------------------------|---------------|-----------|
| Aberdeen         | Glycosyl transferase family 1       | SQI80039.1    | 1,101     |
| Agona            | Hypothetical protein                | ACH51233.1    | 1,284     |
| Albany           | Type II restriction endonuclease subunit R | APV71773.1    | 1,404     |
| Anatum           | Hypothetical protein                | AHW13895.1    | 681       |
| Bareilly         | Hypothetical protein                | QGL28052.1    | 1,113     |
| Berta            | Hypothetical protein                | ESH53086.1    | 522       |
| Blockley         | RNA-dependent DNA polymerase        | EBW652225.2   | 1,193     |
| Braenderup       | Hypothetical protein                | ASO31941.1    | 384       |
| Brandenburg      | Restriction endonuclease subunit S  | KNN28888.1    | 642       |
| Cerro            | Hypothetical protein                | ALI12694.1    | 1,008     |
| Choleraesuis     | LD-carboxypeptidase A               | EFW26856.1    | 966       |
| Corvallis        | Hypothetical protein                | AWD66828.1    | 990       |
| Derby            | Replica family protein             | KMM40244.1    | 846       |
| Dublin           | Conserved hypothetical protein      | ACH74104.1    | 1,161     |
| Elisabethville   | Hypothetical protein                | EBS417135.1   | 1,116     |
| Enteritidis      | Putative phage membrane protein     | CAR32961.1    | 537       |
| Gallinarum       | Hypothetical protein                | AKW12500.1    | 1,020     |
| Give             | DUF1269 domain-containing protein   | OZU63064.1    | 1,662     |
| Hadar            | Hypothetical protein                | KKD69635.1    | 507       |
| Heidelberg       | Abortive infection protein G        | QGF81602.1    | 1,722     |
| I 4,[5],12:i:-   | Hypothetical protein                | QX33576.1     | 231       |
| Infantis         | USG protein                         | CEI4307.1     | 873       |
| Jawana           | Hypothetical protein                | QDI87568.1    | 714       |
| Kedougou         | FRG domain-containing protein       | EBU9135186.1  | 1,410     |
| Kentucky         | Hypothetical protein                | ASQ54323.1    | 1,488     |
| Kottbus          | Phage tail protein                  | EDL01406.1    | 1,479     |
| Litchfield       | Hypothetical protein                | KNL82901.1    | 1,110     |
| Livingstone      | ABC transporter permease subunit    | HAS79095.1    | 909       |
| London           | Hypothetical protein                | ESJ48773.1    | 536       |
| Manhattan        | Reverse transcriptase               | ASO47481.1    | 2,019     |
| Mbandaka         | Hypothetical protein                | AYP83194.1    | 1,824     |
| Meleagridis      | Hypothetical protein                | TSE72805.1    | 2,961     |
| Menston          | DUF4238 domain-containing protein   | ECU3796773.1  | 918       |
| Minnesota        | Amylovoran biosynthesis protein ArniE | APV92684.1    | 810       |
| Mississippi      | AAA family ATPases                  | EDN526886.8   | 4,248     |
| Montevideo       | Hypothetical protein                | AHW10654.1    | 1,767     |
| Muenchen         | Hypothetical protein                | QGH10722.1    | 810       |
| Munster          | Hypothetical protein                | AUM47824.1    | 1,365     |
| Newtoning        | Hypothetical protein                | ECI373925.1   | 195       |
| Newport          | Hypothetical protein                | ALP89662.1    | 843       |
| Ohio             | Adenosylhomocysteinase              | AXE1015.1     | 1,152     |
| Oranienburg      | Hypothetical protein                | AUM45051.1    | 1,392     |
| Panama           | Hypothetical protein                | AKW08656.1    | 1,194     |
| Paratyphi B      | Putative bacteriophage protein      | ESF91711.1    | 1,200     |

(Continued)

### Table 2 | (Continued) Information of gene markers obtained pangenome analysis.

| Serovar          | Protein name                        | Accession no. | Size (bp) |
|------------------|-------------------------------------|---------------|-----------|
| Poona            | Glycosyltransferase                 | SQJ10189.1    | 1,017     |
| Reading          | Hypothetical protein                | KNL67516.1    | 681       |
| Rissen           | Helicase domain-containing protein  | ELX22432.1    | 2,502     |
| Saintpaul        | Hypothetical protein                | AOS37389.1    | 2,277     |
| Schwarzengrund   | YibN protein                        | ACF81479.1    | 2,445     |
| Senftenberg      | FRG domain                         | CRY85796.1    | 822       |
| Singapore        | DNA adenine methylase              | EDA1367050.1  | 673       |
| Stanley          | Hypothetical protein                | ASO28938.1    | 2,142     |
| Tennessee        | Hypothetical protein               | AMW53127.1    | 306       |
| Thompson         | Hypothetical protein               | AGX13681.1    | 1,779     |
| Typhi            | Host-nuclease inhibitor protein     | ALG16175.1    | 531       |
| Typhimurium      | Putative outer membrane            | AAL23519.1    | 558       |
| Uganda           | BREX-1 system phosphatase           | TSB75053.1    | 2,604     |
| Vinchow          | Haloacid dehalogenase-like hydrulase| ECE8801113.1  | 684       |
| Weltevreden      | Putative phosphatase               | CUS01310.1    | 648       |

The specificity of gene markers was evaluated using 535 genomes by in silico analysis. As a result, gene markers were present in the genomes of most target serovars (Figure 2). Notably, most of the gene markers shared 99–100% of the sequence identified in the target serovars, and 0–50% of the sequence identified against other serovars. In contrast, for the serovar Enteritidis, 60 genomes showed 99–100% identity, but in one genome, the Enteritidis gene marker was not found. Among them, Enteritidis 92-0392 contained the Typhimurium gene marker (100% identity) instead of the Enteritidis gene marker, and this genome was determined as Typhimurium in the pangenome analysis and in silico serotyping. For the Typhimurium gene marker, 46 genomes showed 99–100% identity, but in three genomes, the Typhimurium-specific gene marker could not be found. Typhimurium TW-Sm6, FORC50, and FORC98 showed 100% identity to I 4,[5],12:i:-, Enteritidis, and Tennessee-specific gene markers instead of the Typhimurium-specific gene marker, respectively. For serovar Paratyphi B, two genomes showed 100% identity, but in one genome, the Paratyphi B-specific gene marker could not be found. Paratyphi B SARA61 showed 100% identity to Agona gene markers instead of the Paratyphi B gene marker. As mentioned above, misclassified genomes were reclassified. Therefore, Agona, I 4,[5],12:i:-, and Tennessee had more analyzed genomes than the number of corresponding gene markers, so some serovar had more analyzed genomes than the number of corresponding gene markers. Based on the pangenome analysis, serovar-specific
FIGURE 2 | Evaluation of specificity unique gene markers. The figure shows the presence or absence of unique gene markers in 535 genomes. The number of analyzed genomes for each serovar is shown in the green bar, and the number of genomes with gene marker corresponding to that serovar is shown in the red bar.

TABLE 3 | Sequences of serovar-specific primer pairs for real-time PCR.

| Serovar     | Name       | Primer sequences (5′ → 3′)                      | Size (bp) |
|-------------|------------|-------------------------------------------------|-----------|
| Aberdeen    | Aberdeen-F | AAC AAG GGG TAC AGG GAT TA                      | 131       |
|             | Aberdeen-R | AAT CCT TAT TAT GCG CCC CA                      |           |
| Agona       | Agona-F    | GCA TCT GGC GGC AAG TCA TA                      | 190       |
|             | Agona-R    | GTG AGC GTA ATG GGG ATG TA                      |           |
| Albany      | Albany-F   | TAG TCA GTG AGC ACC GAG TT                      | 163       |
|             | Albany-R   | ACG CCA TGT AGA TTC GTT AT                      |           |
| Anatum      | Anatum-F   | AAA GCA CCC TGA GTC AGA TG                      | 143       |
|             | Anatum-R   | CAA GTC CAC CGA CTG CTA                         |           |
| Bareilly    | Bareilly-F | GGT GGT AGT AAC AAG GAT TG                      | 114       |
|             | Bareilly-R | ACT GCT AGT TCC GTC GTA AG                      |           |
| Berta       | Berta-F    | TAA CGG AGG CAA CAA CAG TG                      | 145       |
|             | Berta-R    | CGG AGA GGG TCC AGT TGT TT                      |           |
| Blockley    | Blockley-F | TCC GTG TGT CAT GAG CAG TT                      | 124       |
|             | Blockley-R | GAA GAT CAT CAC GCC TAG GT                      |           |
| Braenderup  | Braenderup-F| GGA GAA TGC TCG CAG GAA GA                     | 156       |
|             | Braenderup-R| GCT GGT TCA AAG TAA TGC GG                     |           |
| Brandenburg | Brandenburg-F| TCG TTC TAC ACC TAA AGG TGG                    | 108       |
|             | Brandenburg-R| TAC GCG ACA TCA TGT ACC CT                     |           |
| Cerro       | Cerro-F    | CGT TTC TCC GAT TAT GGA GC                     | 157       |
|             | Cerro-R    | GGC ATT GTT ACA GAC AGC AA                      |           |
| Choleraesuis| Choleraesuis-F | GCT CCA TCT GCG CAA ATT GT                     | 125       |
|             | Choleraesuis-R | TCC AGG AAC GCT GTA GGC TCG                   |           |
| Corvallis   | Corvallis-F | AAG CGT TTA TTT GAG GCT GA                     | 139       |
|             | Corvallis-R | CGC TGT TCG ATG CTT CAA GT                     |           |
| Derby       | Derby-F    | TGC GTC GTT TCA TCA ATG TG                     | 106       |
|             | Derby-R    | CGT TGT AGC OCA TCT AGG TT                     |           |
| Dublin      | Dublin-F   | GCG TCA AGG TTT ATT TAA CAG GG                 | 114       |
|             | Dublin-R   | GGA TGT CAA TCG TGT GC                        |           |
| Elisabethville| Elisabethville-F | GAC CAC GAC CGG TAC AGC AA                   | 178       |
|             | Elisabethville-R | CGG CGG ATA CTA CAG CAC AGA                     |           |
| Enteritidis | Enteritidis-F | TTG GTA AAT CCG TCG GAC AA                   | 105       |
|             | Enteritidis-R | AAT CGG TAC CGG CCT CAA TA                     |           |
| Gallinarum  | Gallinarum-F | CGA CGG TCG TCA ATC CTA CT                     | 113       |
|             | Gallinarum-R | ATC AAC CAC AGC CGT AGC AG                     |           |
| Give        | Give-F     | TCA TTG GCA CGG GTG AGT CG                      | 103       |
|             | Give-R     | CCT TGC TAT CAC GGC ACA TC                     |           |
| Hadar       | Hadar-F    | GTG AGT CTT TTT CCG TGA TA                     | 162       |
|             | Hadar-R    | ATC TCA CCC ATT CAC AGA TA                     |           |
| Heidelberg  | Heidelberg-F | CGG CGA ATT AAT CAT ACG AG                     | 105       |
|             | Heidelberg-R | CTC TCA CCT GAT TTT GCC TGT                     |           |
| I 4,[5],12:i-| I 4,[5],12:i-F | AAG TGC GGC AGT TAG CCT CT                   | 113       |
|             | I 4,[5],12:i-R | GGT ATC GCC GTC AAT ACA CA                     |           |
| Infantis    | Infantis-F | GTG CGA GAT GGG TAT GTG QC                      | 109       |
|             | Infantis-R | CAG GAG TTC CGG CAC AGA                       |           |
| Javiana     | Javiana-F  | TGG CTA CTC AGG CAG TAC TA                     | 112       |
|             | Javiana-R  | AGC ATA ACT CGG TGA TT                        |           |
| Kedougou    | Kedougou-F | GTC AGG CCT CTT GAA GAT GT                     | 102       |
|             | Kedougou-R | CTC CGG TTC ATA CAA TCG TG                     |           |
| Kentucky    | Kentucky-F | CGA TAA CCT GCG GAG TCA GA                     | 141       |
|             | Kentucky-R | TTC CAC CGG TCG CCT CAT AA                     |           |
| Kottbus     | Kottbus-F  | GCG TCT GAC TGG AGC AGA TT                     | 108       |
|             | Kottbus-R  | ACC ACA GTC AAC GCC TAG GT                     |           |
It is crucial to develop the primer pair with high specificity as sensitivity and specificity (Hoorfar, 2011; Xiong et al., 2018). Many isolates with relatively simple procedures; it also has potent contrast, the PCR-based method can rapidly detect pathogenic bacteria with high accuracy and sensitivity (Abubakar et al., 2007; Chen et al., 2010). This method can cost-effectively detect additional bioinformatics analysis (Ibrahim and Morin, 2018). In this study, our method was applied against 232 Salmonella strains. However, since the number of strains included in some serovars (e.g., Aberdeen, Berta, Cerro, Hadar, Kedougou, etc.) is rarely isolated, only a small number of strains were analyzed. In silico PCR was performed using the 625 genomes to determine whether marker genes are representative of each serovar and whether the accuracy can maintain high enough once it is applied to more strains of these serovars. All primer pairs that can accurately detect Salmonella serovars were developed (Table 3). An Infantis-specific primer pair was developed in the previous study (Yang et al., 2021).

| Serovar | Name | Primer sequences (5′→3′) | Size (bp) |
|---------|------|--------------------------|----------|
| Litchfield | Litchfield-F | CAG ACT TAA TAG AGG ACC CA | 143 |
| Litchfield-R | GTA TCC ATT GGC AGC GAG TA | 102 |
| Livingstone | Livingstone-F | CTC GTG TTT ATT CCA TCA AC | 103 |
| Livingstone-R | ACA AAG CCT TCA GGT GG | 126 |
| London | London-F | GTC TCA TCC GGA AGC AAC AA | 141 |
| London-R | TGT GCC ATT CTC TGG CTC AT | 118 |
| Manhattan | Manhattan-F | GCT GAT GCA GGG TAG CAA TA | 128 |
| Manhattan-R | GTA TCC ATT GGC AGC GAG TAA | 126 |
| Mbandaka | Mbandaka-F | TTC GGA TTA GAA GGC ATGCT C | 130 |
| Mbandaka-R | GCC GCA ATG CCA CCA AGA AA | 126 |
| Meleagridis | Meleagridis-F | TGG CGA TAT ACC GGG TAG CT | 118 |
| Meleagridis-R | CAG GTG TCG AAA GGG TAG ATG | 126 |
| Menston | Menston-F | TAG TGG TGC GAC GGA GGT GA | 101 |
| Menston-R | GCC GCA ATG CCA CCA AGA AA | 126 |
| Minnesota | Minnesota-F | TCC GCG TAA CCG ATG ACT AT | 117 |
| Minnesota-R | TCC GGA TAT CCG GAA TGG TCC | 117 |
| Mississippi | Mississippi-F | CCA CGA CAT CAT TCA TC | 115 |
| Mississippi-R | GCC ATG GGC ACT AAT GA | 126 |
| Montevideo | Montevideo-F | CCA ACC TGG CCA ACA AGA TC | 120 |
| Montevideo-R | U bacteria-F | ATG TGG TAA TGG ACT GG | 265-267 |
| Muenchen | Muenchen-F | GCA CGT ATG CAG ATC GAA | 134 |
| Muenchen-R | GTA TCC ATT GGC AGC GAG TA | 102 |
| Muenster | Muenster-F | CAC CTG CCT GGC CGG ATT GAT | 175 |
| Muenster-R | GCC GCA ATG CCA CCA AGA AA | 126 |
| Newtoning | Newtoning-F | CGT AGT CGT GGT TGG TGG TA | 121 |
| Newtoning-R | GCC GCA ATG CCA CCA AGA AA | 126 |
| Newport | Newport-F | GTT GCC AAA AAG CAC AAT GA | 117 |
| Newport-R | GCC GCA ATG CCA CCA AGA AA | 126 |
| Ohio | Ohio-F | CGA TAA TTG CCG CCT TCT GTA | 104 |
| Ohio-R | GCC GCA ATG CCA CCA AGA AA | 126 |
| Oranienburg | Oranienburg-F | GGT GAT ATT GGG ATT CCA CC | 101 |
| Oranienburg-R | GCC GCA ATG CCA CCA AGA AA | 126 |
| Panama | Panama-F | GCC TCA TTA GAT CCA ACA GC | 104 |
| Panama-R | GCC GCA ATG CCA CCA AGA AA | 126 |
| Paratyphi B | Paratyphi B-F | CGA TGG CTC CTT GCT ATT CCA G | 120 |
| Paratyphi B-R | GCC GCA ATG CCA CCA AGA AA | 126 |
| Poona | Poona-F | GCT GGG CGG ACA ATC ATC AAC C | 120 |
| Poona-R | GCC GCA ATG CCA CCA AGA AA | 126 |
| Reading | Reading-F | GCA ATC GAC GTC AGT CGT GTA | 127 |
| Reading-R | GCC GCA ATG CCA CCA AGA AA | 126 |
| Rissen | Rissen-F | GGT AGC ATT GGC AAG TGA | 117 |
| Rissen-R | GCC GCA ATG CCA CCA AGA AA | 126 |
| Saintpal | Saintpal-F | TGG GTA GAA GTG CCT GGA | 154 |
| Saintpal-R | GCC GCA ATG CCA CCA AGA AA | 126 |
| Schwarzengrund | Schwarzengrund-F | CAA TGG CGG ACC ACC ACC TA | 131 |
| Schwarzengrund-R | GCC GCA ATG CCA CCA AGA AA | 126 |
| Senttenberg | Senttenberg-F | GAG CAG GAT ATG CGC GAC TA | 131 |
| Senttenberg-R | GCC GCA ATG CCA CCA AGA AA | 126 |
| Singapore | Singapore-F | CCA GTA GAG AAT AAC ATG G | 109 |
| Singapore-R | GCC GCA ATG CCA CCA AGA AA | 126 |
| Stanley | Stanley-F | CAG ACC TGG GAT TTA GGT TA | 120 |
| Stanley-R | GCC GCA ATG CCA CCA AGA AA | 126 |

1 Internal standard primer pair was developed by Aboutalebian et al. (2021).

**Specificity and Accuracy of the Developed Real-Time Polymerase Chain Reaction**

Pangenome analysis based on the whole-genome sequence can efficiently select serovar-specific gene markers using large-scale genomes. The classification of pathogenic bacteria to their correct taxonomy using whole-genome sequencing shows reproducibility and accuracy, but is expensive and requires additional bioinformatics analysis (Ibrahim and Morin, 2018). In contrast, the PCR-based method can rapidly detect pathogenic bacteria with high accuracy and sensitivity (Abubakar et al., 2007; Chen et al., 2010). This method can cost-effectively detect many isolates with relatively simple procedures; it also has potent sensitivity and specificity (Hoorfar, 2011; Xiong et al., 2018). It is crucial to develop the primer pair with high specificity as
the accuracy of these PCR-based assays depends mainly on the specific gene or primer pair. Therefore, in this study, a real-time PCR method was developed for serotyping by detecting unique gene markers obtained through pangenome analysis.

A total of 199 *Salmonella* strains and 29 non-*Salmonella* strains were used to develop a real-time PCR method that is specific and accurate. Amplification plots for the most frequently isolated six serovars worldwide are shown in Figure 3, and the
TABLE 4 | Comparison of serotyping results through antisera aggregation and real-time PCR.

| Serovar (no. of strains) | Antigenic formulas¹ | No. of amplified strains² |
|-------------------------|---------------------|--------------------------|
| Aberdeen (n = 2)        | 11,12:1,2            | 2                        |
| Agona (n = 4)           | 1,4,5,12:1,2,12,1,7  | 4                        |
| Albany (n = 1)          | 8,20:z1,2           | 1                        |
| Anatum (n = 2)          | 3,10,15:15,34:e,1:6  | 2                        |
| Bareilly (n = 8)        | 6,7,14:y:1,5        | 8                        |
| Berta (n = 1)           | 1,9,12:f,g,b,-       | 1                        |
| Blockley (n = 2)        | 8:k,1,5             | 2                        |
| Braenderup (n = 4)      | 6,7,14:e,n,z15b     | 4                        |
| Brandenburg (n = 2)     | 4,5,12:i,v,n,z15b   | 2                        |
| Cero (n = 2)            | 6,14,18:z23,1:1,5   | 2                        |
| Choleraesuis (n = 1)    | 7,c,1,5             | 1                        |
| Corvallis (n = 1)       | 8,20,z23,23 [69]    | 1                        |
| Derby (n = 5)           | 1,4,5,12:1,2        | 5                        |
| Dublin (n = 2)          | 1,9,12:g,p,-        | 2                        |
| Elisabethhile (n = 1)   | 3,[10]:15:r,1,7     | 21                       |
| Enteritidis (n = 21)    | 1,9,12:g,n,-        | 21                       |
| Gallinarum (n = 1)      | 1,9,12:1           | 1                        |
| Give (n = 3)            | 3,10,15:15,34:i:v:1,7 | 2             |
| Hadar (n = 7)           | 3,[10]:e,n,x        | 7                        |
| Heidelberg (n = 4)      | 1,4,5,12:i,z:1,2    | 4                        |
| I 4,5,12:12:12:1:2      | 2,4,[5],12:i,1      | 2                        |
| Infantis (n = 8)        | 8,7,14:r,1,5        | 8                        |
| Javanila (n = 1)        | 1,9,12:z23:1,5      | 1                        |
| Kedougou (n = 2)        | 1,13,23:i,l,w       | 2                        |
| Kentucky (n = 2)        | 8,20:z23            | 2                        |
| Kottbus (n = 1)         | 8:e,h,1,5           | 1                        |
| Litchfield (n = 2)      | 8:l,v:1,2           | 2                        |
| Livingstone (n = 5)     | 6,7,14:d,l,w        | 5                        |
| London (n = 2)          | 3,[10]:15:i,v:1,6   | 2                        |
| Manhattan (n = 2)       | 6,8:d,1,5           | 2                        |
| Mbandaka (n = 2)        | 6,7,14:z10:e,n,z15  | 2                        |
| Meleagridis (n = 1)     | 3,10,15:15,34:e,h,l,w | 1             |
| Menston (n = 1)         | 6,7:1,4,6[b],1,5    | 1                        |
| Minnesota (n = 1)       | 21:e,n,x            | 1                        |
| Mississippi (n = 1)     | 1,13,23:b,1,5       | 1                        |
| Montevideo (n = 8)      | 8,7,14:g,m,p,s,1,2,7 | 8                        |
| Muenchen (n = 5)        | 6,8:d,1,2           | 5                        |
| Muenster (n = 1)        | 3,[10]:15,34:e,h,1,5 | 1                        |
| Newtoning (n = 2)       | 3,[10]:15,34:e,h,1,6 | 2                        |
| Newport (n = 3)         | 8:e,h,1,2           | 3                        |
| Ohio (n = 4)            | 6,7,14:b,l,w        | 4                        |
| Oranienburg (n = 2)     | 6,7,14:m,t,[z23]    | 2                        |
| Panama (n = 3)          | 1,9,12:v,1,5        | 3                        |
| Paratyphi B (n = 2)     | 1,4,5,12:b,1,2      | 2                        |
| Poona (n = 2)           | 1,13,22:z,1,6       | 2                        |
| Reading (n = 3)         | 1,4,5,12:e,h,1,5    | 3                        |
| Rissen (n = 3)          | 8,7,14:g,f,-        | 3                        |
| Saintpaul (n = 3)       | 1,4,5,12:e,h,1,2    | 3                        |
| Schwarzengrund (n = 3)  | 1,4,12:27,d,1,7     | 3                        |
| Sentfenten (n = 4)      | 1,3,19:g,t,1,-      | 4                        |
| Singapore (n = 1)       | 6,7:k,e,n,x         | 1                        |
| Stanley (n = 3)         | 1,4,5,12:27:d,1,2   | 3                        |
| Tennessee (n = 2)       | 6,7,14:z23,1,2,7    | 2                        |

(Continued)

TABLE 4 | (Continued)

| Serovar (no. of strains) | Antigenic formulas¹ | No. of amplified strains² |
|-------------------------|---------------------|--------------------------|
| Thompson (n = 5)        | 6,7,14:k,1,5        | 5                        |
| Typhii (n = 5)          | 9,12:cd,-           | 5                        |
| Typhimurium (n = 8)     | 1,4,5,12:i,1,2      | 8                        |
| Uganda (n = 1)          | 3,[10]:15,i,z:1,5   | 1                        |
| Vinchowdy (n = 1)       | 28:mt:e,n,z15       | 1                        |
| Virchow (n = 8)         | 6,7,14:r,1,2        | 8                        |
| Weltevreden (n = 1)     | 3,[10]:15,v,z6      | 1                        |

¹Serovar confirmed through antisera aggregation.
²Number of strains amplified through real-time PCR with corresponding primer pair.

TABLE 5 | Serotyping results for isolates with unknown serovars.

| Strains | Strain designations | Real-time PCR |
|---------|---------------------|---------------|
| KVCC1-BV1900349 | Salmonella enterica | I 4,5,12:i,- |
| KVCC-BV1700172 | Salmonella enterica | Entertidis |
| KVCC-BV1700171 | Salmonella enterica | Entertidis |
| KVCC-BV1700170 | Salmonella enterica | Entertidis |
| KVCC-BV1700168 | Salmonella enterica | Entertidis |
| KVCC-BV1700169 | Salmonella enterica | Entertidis |
| KVCC-BV0001439 | Salmonella enterica | Blockley |
| KVCC-BV000429  | Salmonella enterica | Blockley |
| KVCC-BV1800008 | Salmonella enterica | Hadar |
| KVCC-BV000718  | Salmonella enterica | Livingstone |
| KVCC-BV000704  | Salmonella enterica | Livingstone |
| KVCC-BV1800002 | Salmonella enterica | Mbandaka |
| KVCC-BV1800001 | Salmonella enterica | Mbandaka |
| KVCC-BV1800010 | Salmonella enterica | Montevideo |
| KVCC-BV1900334 | Salmonella enterica | Rissen |
| KVCC-BV0701417 | Salmonella enterica | Typhimurium |
| KVCC-BV000703  | Salmonella enterica | Typhimurium |
| KVCC-BV000695  | Salmonella enterica | Typhimurium |
| KVCC-BV000692  | Salmonella enterica | Typhimurium |
| KVCC-BV000691  | Salmonella enterica | Typhimurium |
| KVCC-BV000702  | Salmonella enterica | Typhimurium |
| KVCC-BV000699  | Salmonella enterica | Typhimurium |
| KVCC-BV000683  | Salmonella enterica | Typhimurium |
| KVCC-BV000689  | Salmonella enterica | Typhimurium |
| KVCC-BV1300166 | Salmonella species  | Enteridis |
| KVCC-BV1300261 | Salmonella species  | Infantis |
| KVCC-BV1300155 | Salmonella species  | Infantis |
| KVCC-BV1300151 | Salmonella species  | Montevideo |
| KVCC-BV1300258 | Salmonella species  | Typhimurium |
| KVCC-BV1600002 | Salmonella species  | Virchow |
| KVCC-BV1200044 | Salmonella species  | Infantis |
| KVCC-BV1800599 | Salmonella species  | London |
| KVCC-BV1800010 | Salmonella species  | Senftenberg |

¹KVCC, the Korea Veterinary Culture Collection.

result on the remaining serovars is in Supplementary Table 6. The genomic DNA across serovars yielded a detectable amplicon for the target primer pair, whereas those from all non-target Salmonella did not generate any signal (Figure 3). The Ct
value ranges were 11.49–18.07 for each *Salmonella* serovar strain (Supplementary Table 6). Thus, primer pairs developed in this study were considered specific for the identification and detection of individual *Salmonella* serovars. Serial dilution was used on the genomic DNA of *Salmonella* reference strains to confirm the accuracy of the real-time PCR assay. All *Salmonella* serovar-specific primer pairs showed a linear relationship over the range of 0.002–20 ng. The slopes for the primer pairs of Bareilly, Enteritidis, I 4,[5],12:i:−, Montevideo, Typhi, and Typhimurium were $-3.589, -3.395, -3.66, -3.457, -3.677$, and $-3.61$, respectively, and the $R^2$ values were $≥0.997$ (Figure 4). The primer pairs for the remaining 54 serovars also showed that slope values were $-3.19$ to $-3.683$, and the $R^2$ values were $≥0.996$ (Supplementary Table 7). To show high efficiency, the slope value should be $-3.1$ to $-3.9$, and the $R^2$ value $≥0.996$ for the standard curve (Broeders et al., 2014). The slope and $R^2$ values of all primer pairs developed in this study were within these ranges.

### Evaluation of Real-Time Polymerase Chain Reaction and Validation by Traditional Serotyping

The real-time PCR method developed in this study was evaluated in a single 96-well plate using 222 *Salmonella* strains. Moreover, serovars of some strains were identified using the antisera agglutination method and compared with real-time PCR results. Real-time PCR competency was checked using an internal standard. Of 222 strains, 189 strains are known at the serovar level, and the remaining 33 were strains with unknown serovars. The real-time PCR result determined that the corresponding serovars were detected when amplified in a well containing a serovar-specific primer pair. All strains were amplified in wells containing a specific primer pair and internal standard primer pair, whereas no amplification was observed in the other wells (Supplementary Table 8). The internal standard primer pair was amplified in 222 *Salmonella* strains and 29 non-*Salmonella*, and the Ct value ranged from 10.03 to 14.95. Serovars of all strains were identified accurately by real-time PCR, and the result was identical to the serotyping result of the antisera agglutination method (Table 4). Anatum and Newington presented similar antigenic formulas as the result of antisera agglutination, but two serovars were distinguishable in real-time PCR. To verify our real-time PCR method, serovars were determined for 33 isolates identified down to the genus or species level obtained from the Korea Veterinary Culture Collection (KVCC) (Table 5). As a result, 33 isolates were determined as 13 different serovars, such as I 4,[5],12:i:−, Enteritidis, Blockley, Hadar, Livingstone, Mbandaka, Montevideo, Rissen, Typhimurium, Infantis, Virchow, London, and Senftenberg. Therefore, our results suggest that the 60 primer pairs and real-time PCR method developed in this study are 100% accurate in detecting *Salmonella* serovars. However, in this study, few strains were used in some serovars that may influence the identification of the real marker genes and detecting accuracy.

This method not only can clearly distinguish between two serovars presenting similar antigenic formulas, but also alleviates the time and cost required for traditional serotyping method. This method has the limitation of lack of replicates in a sample run of 60 serovars in 96-well plates, but may be necessary for application in the field as it is evaluated in a single plate.

### CONCLUSION

In this study, novel serovar-specific gene markers were discovered through pangenome analysis of whole-genome sequences. The pangenome analysis could identify gene markers for 60 *Salmonella* serovars present in genomes of target serovars and absent in genomes of other serovars. Furthermore, *in silico* analyses confirmed that some genomes deposited in the public database, such as the NCBI, were incorrectly designated. The real-time PCR method, designed to detect serovar-specific gene markers using a single 96-well plate, successfully detected 222 strains, thus validating the specificity and effectiveness of the assay. Additionally, the traditional serotyping method yielded ambiguous results for strains that share similar antigenic formulas but were accurately identified as one serovar using real-time PCR. These results suggest that the efficient real-time PCR assay developed could be used as a high-throughput diagnostic tool to identify 60 serovars. The real-time PCR method developed in this study is useful in diagnosing *Salmonella* infections and has applications in food safety and human health.

### DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author/s.

### AUTHOR CONTRIBUTIONS

S-MY and H-YK contributed to the conception and design of this study. S-MY, EK, SK, and DoK performed the analysis of *in silico* serotyping and pangenome. S-MY, EK, DaK, and H-BK performed the real-time PCR. JB and HY performed the serotyping using antisera agglutination. S-MY and EK prepared a draft manuscript. H-YK reviewed and edited the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version.

### FUNDING

This research was supported by a grant (19162MFDS042) from Ministry of Food and Drug Safety in 2021.

### SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.750379/full#supplementary-material
REFERENCES

Abd-Elsalam, K. A. (2003). Bioinformatic tools and guideline for PCR primer design. Afr. J. Biotechnol. 2, 91–100. doi: 10.5897/abj2003.00010

Aboutalebian, S., Ahmadikia, K., Fakhim, H., Chavazizadeh, J., Okhovat, A., Nikaen, M., et al. (2021). Direct Detection and Identification of the Most Common Bacteria and Fungi Causing Otitis Externa by a Stepwise Multiplex PCR. Front. Cell. Infect. Microbiol. 11:644060. doi: 10.3389/fcimb.2021.64060

Abubakar, I., Irvine, L., Aldus, C. F., Wyatt, G. M., Fordham, R., Schelenz, S., et al. (2007). A systematic review of the clinical, public health and cost-effectiveness of rapid diagnostic tests for the detection and identification of bacterial intestinal pathogens in faeces and food. Health Technol. Assess. 11, 1–216. doi: 10.3310/hta11360

Akiba, M., Kusumoto, M., and Iwata, T. (2011). Rapid identification of Salmonella enterica serovars, Typhimurium, Choleraesuis, Infantis, Hadar, Enteritidis, Dublin and Gallinarum, by multiplex PCR. J. Microbiol. Methods 85, 9–15. doi: 10.1016/j.mimet.2011.02.002

Broeders, S., Huber, I., Grohmann, L., Berben, G., Taverniers, I., Mazzara, M., et al. (2014). Guidelines for validation of qualitative real-time PCR methods. Trends Food Sci. Technol. 37, 115–126. doi: 10.1016/j.tifs.2014.03.008

Chaudhari, N. M., Gupta, V. K., and Dutta, C. (2016). BPGA-an ultra-fast pan-genome analysis pipeline. Sci. Rep. 6:24373. doi: 10.1038/srep24373

Chen, J., Zhang, L., Paoli, G. C., Shi, C., Tu, S. I., and Shi, X. (2010). A real-time PCR method for the detection of Salmonella enterica from food using a target sequence identified by comparative genomic analysis. Int. J. Food Microbiol. 137, 168–174. doi: 10.1016/j.ijfoodmicro.2009.12.004

Cooper, A. L., Low, A. J., Koziol, A. G., Thomas, M. C., Leclair, D., Tamber, S., et al. (2020). Systematic Evaluation of Whole Genome Sequence-Based Predictions of Salmonella Serotype and Antimicrobial Resistance. Front. Microbiol. 11:549. doi: 10.3389/fmicb.2020.00549

Cos, S., Brown, L., and Green, G. (2021). Point and Pair: A Multiplex PCR platform with pairwise probe comparison. Front. Microbiol. 12:750379. doi: 10.3389/fmicb.2021.750379

Dietrich, Y., Tennekoon, R., Hwang, R. W., Lee, K. H., and Qin, X. (2012). Rapid Differentiation of Salmonella enterica Serovars, Typhimurium and Enteritidis by a Multiplex PCR Assay. Front. Microbiol. 3, 287–93. doi: 10.3389/fmicb.2012.02877

Draper, S. J., Vis, M. L., and Moolenaar, R. (2020). Rapid Identification and Differentiation of Salmonella enterica subsp. enterica Serovars by Multiplex PCR. J. Microbiol. Methods 180:106181. doi: 10.1016/j.mimet.2020.106181

Eaben, R. M., Esem, O. 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. doi: 10.7717/peerj.1319

Ehret, G. A., Arakawa, T., and Himeno, T. (2013). Development of a rapid multiplex PCR assay for the detection of Salmonella enterica subsp. enterica serovars. Int. J. Food Microbiol. 161, 259–268. doi: 10.1016/j.ijfoodmicro.2013.02.014

Ehret, G. A., and Arakawa, T. (2014). Development of a rapid and specific multiplex PCR assay for the detection of Salmonella enterica subsp. enterica serovars. J. Microbiol. Methods 101, 1–8. doi: 10.1016/j.mimet.2013.07.010

Ehret, G. A., Arakawa, T., and Himeno, T. (2015). Rapid and specific multiplex PCR assay for the detection of Salmonella enterica subsp. enterica serovars. J. Microbiol. Methods 113, 23–28. doi: 10.1016/j.mimet.2015.03.022

Ehret, G. A., Arakawa, T., and Himeno, T. (2016). Development of a rapid multiplex PCR assay for the detection of Salmonella enterica subsp. enterica serovars. Front. Microbiol. 7, 1209–1219. doi: 10.3389/fmicb.2016.01209

Ehret, G. A., Arakawa, T., and Himeno, T. (2017). Development of a rapid and specific multiplex PCR assay for the detection of Salmonella enterica subsp. enterica serovars. Front. Microbiol. 8, 1978–87. doi: 10.3389/fmicb.2017.01978

Ehret, G. A., Arakawa, T., and Himeno, T. (2018). Development of a rapid and specific multiplex PCR assay for the detection of Salmonella enterica subsp. enterica serovars. Front. Microbiol. 9, 1978–87. doi: 10.3389/fmicb.2018.01978

Ehret, G. A., Arakawa, T., and Himeno, T. (2019). Development of a rapid and specific multiplex PCR assay for the detection of Salmonella enterica subsp. enterica serovars. Front. Microbiol. 10, 108566. doi: 10.3389/fmicb.2019.108566

Ehret, G. A., Arakawa, T., and Himeno, T. (2020). Development of a rapid and specific multiplex PCR assay for the detection of Salmonella enterica subsp. enterica serovars. Front. Microbiol. 11, 601024. doi: 10.3389/fmicb.2020.601024

Ehret, G. A., Arakawa, T., and Himeno, T. (2021). Development of a rapid and specific multiplex PCR assay for the detection of Salmonella enterica subsp. enterica serovars. Front. Microbiol. 12, 750379. doi: 10.3389/fmicb.2021.750379

Ehret, G. A., Arakawa, T., and Himeno, T. (2022). Development of a rapid and specific multiplex PCR assay for the detection of Salmonella enterica subsp. enterica serovars. Front. Microbiol. 13, 4060. doi: 10.3389/fmicb.2022.4060

Ehret, G. A., Arakawa, T., and Himeno, T. (2023). Development of a rapid and specific multiplex PCR assay for the detection of Salmonella enterica subsp. enterica serovars. Front. Microbiol. 14, 750379. doi: 10.3389/fmicb.2023.750379
Ye, Q., Shang, Y., Chen, M., Pang, R., Li, F., Xiang, X., et al. (2021). Identification of Novel Sensitive and Reliable Serovar-Specific Targets for PCR Detection of Salmonella Serovars Hadar and Albany by Pan-Genome Analysis. *Front. Microbiol.* 12:605984. doi: 10.3389/fmicb.2021.605984

Yoshida, C. E., Kruczkiewicz, P., Laing, C. R., Lingohr, E. J., Gannon, V. P. J., Nash, J. H. E., et al. (2016). The *Salmonella* in silico typing resource (SISTR): an open web-accessible tool for rapidly typing and subtyping draft *Salmonella* genome assemblies. *PLoS One* 11:e0147101. doi: 10.1371/journal.pone.0147101

Zhang, S., den Bakker, H. C., Li, S., Chen, J., Dinsmore, B. A., Lane, C., et al. (2019). SeqSero2: rapid and improved *Salmonella* serotype determination using whole-genome sequencing data. *Appl. Environ. Microbiol.* 85, e01746–19. doi: 10.1128/AEM.01746-19

Zhang, S., Yin, Y., Jones, M. B., Zhang, Z., Kaiser, B. L. D., Dinsmore, B. A., et al. (2015). *Salmonella* serotype determination utilizing high-throughput genome sequencing data. *J. Clin. Microbiol.* 53, 1685–1692. doi: 10.1128/JCM.00323-15

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Yang, Kim, Kim, Baek, Ko, Kim, Yoon and Kim. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.