



OPEN

Whole genome sequencing and protein structure analyses of target genes for the detection of *Salmonella*

Lijun Hu¹, Guojie Cao¹, Eric W. Brown¹, Marc W. Allard¹, Li M. Ma² & Guodong Zhang¹✉

Rapid and sensitive detection of *Salmonella* is a critical step in routine food quality control, outbreak investigation, and food recalls. Although various genes have been the targets in the design of rapid molecular detection methods for *Salmonella*, there is limited information on the diversity of these target genes at the level of DNA sequence and the encoded protein structures. In this study, we investigated the diversity of ten target genes (*invA*, *fimA*, *phoP*, *spvC*, and *agfA*; *ttrRSBCA* operon including 5 genes) commonly used in the detection and identification of *Salmonella*. To this end, we performed whole genome sequencing of 143 isolates of *Salmonella* serotypes (Enteritidis, Typhimurium, and Heidelberg) obtained from poultry (eggs and chicken). Phylogenetic analysis showed that *Salmonella* ser. Typhimurium was more diverse than either Enteritidis or Heidelberg. Forty-five non-synonymous mutations were identified in the target genes from the 143 isolates, with the two most common mutations as T ↔ C (15 times) and A ↔ G (13 times). The gene *spvC* was primarily present in *Salmonella* ser. Enteritidis isolates and absent from Heidelberg isolates, whereas *ttrR* was more conserved (0 non-synonymous mutations) than *ttrS*, *ttrB*, *ttrC*, and *ttrA* (7, 2, 2, and 7 non-synonymous mutations, respectively). Notably, we found one non-synonymous mutation (*fimA*-Mut.6) across all *Salmonella* ser. Enteritidis and *Salmonella* ser. Heidelberg, C → T (496 nt position), resulting in the change at AA 166 position, Glutamine (Q) → Stop codon (TAG), suggesting that the *fimA* gene has questionable sites as a target for detection. Using Phyre² and SWISS-MODEL software, we predicted the structures of the proteins encoded by some of the target genes, illustrating the positions of these non-synonymous mutations that mainly located on the α-helix and β-sheet which are key elements for maintaining the conformation of proteins. These results will facilitate the development of sensitive molecular detection methods for *Salmonella*.

Foodborne diseases caused by *Salmonella* are an ongoing public health problem and a global economic burden. Non-typhoidal *Salmonella* caused an estimated 1,027,561 illnesses, 19,336 hospitalizations, and 378 deaths annually in the United States¹. In 2010, World Health Organization (WHO) reported that non-typhoidal *Salmonella enterica* was the top causing agent responsible for 59,000 of the 230,000 (25.65%) deaths, and 4 million of the 18 million (22.22%) Disability Adjusted Life Years (DALYs) attributed to diarrheal disease agents². *Salmonella*-contaminated foods, especially poultry-derived foods (eggs, chicken meat), are the most significant source of salmonellosis³. *Salmonella* serotypes Enteritidis, Typhimurium, and Heidelberg are consistently reported as the most frequent *Salmonella* serotypes associated with egg and poultry products^{4–6}. However, global egg and chicken consumptions have been rising steadily; egg and chicken meat production increased from 14.7 to 66.4 million tons (350.2%) and 7.9 to 92.8 million tons (1077.2%), respectively, from 1962 to 2012⁷. Therefore, the routine surveillance, detection, and prevention of *Salmonella* in poultry products are extremely important to public health.

Detection methods are essential for surveillance and prevention of foodborne pathogens. All effective molecular detection methods (such as PCR based methods and isothermal methods) require target genes. The target genes *invA*, *ttrRSBCA* (operon including 5 genes of *ttrR*, *ttrS*, *ttrB*, *ttrC*, *ttrA*), *phoP*, *fimA*, *agfA*, and *spvC* have been frequently used either separately or in combination for *Salmonella* detection in many PCR and

¹Division of Microbiology, Office of Regulatory Science, Center for Food Safety and Nutrition, U.S. Food and Drug Administration, 5001 Campus Dr., College Park, MD 20740, USA. ²Institute for Biosecurity and Microbial Forensics, Department of Entomology and Plant Pathology, Oklahoma State University, Stillwater, OK 74074, USA. ✉email: Guodong.Zhang@FDA.HHS.GOV

loop-mediated isothermal amplification (LAMP) assays^{8–18}. All these target genes are located on chromosome, except for *spvC*, which is encoded on a plasmid. The proteins they encode have a variety of functions and are all involved in virulence, such as participation in the biosynthesis of flagella via an encoded putative inner membrane protein (*invA*), tetrathionate respiration (*ttrRSBCA* operon), regulation and two-component signaling (*phoP*), adherence and fimbriae activity (*fimA* and *agfA*), and promotion of the survival and rapid growth of *Salmonella* in the host (*spvC*). Considering the importance of these genes in the development of molecular detection methods for *Salmonella*, variations in their sequences among *Salmonella* isolates may impact the target coverages of these detection methods. The recent advances in whole-genome sequencing (WGS) technology and bioinformatics provide powerful tools for studying the possible variations in these target genes effectively among large number of *Salmonella* isolates. Furthermore, WGS and bioinformatics are also the most powerful tools to genotype various microorganisms, including *Salmonella*. There are WGSs from 155,509 *Salmonella enterica* isolates provided online²⁰ and these data have been widely used to track and investigate outbreaks. For example, WGS was used by Inns and colleagues²¹ to investigate 136 *Salmonella* ser. Enteritidis cases from an outbreak in the United Kingdom in 2015, including 21 travel-associated cases of salmonellosis. That outbreak was traced back to chicken eggs and further linked to 18 contemporaneous cases reported in Spain (18 identified cases). In another similar study, WGS data and food trace-back investigations identified that eggs used at food premises were the sources of *Salmonella* ser. Typhimurium contamination in seven outbreaks in Australia, with WGS technology providing a higher discriminatory ability than multiple-locus variable-number tandem repeat analysis (MLVA)²².

As serotypes of Enteritidis, Typhimurium, and Heidelberg are the common causes of salmonellosis worldwide, we chose to study a group of representative isolates of these serotypes sourced from poultry. The purposes of this study were to (1) determine the sequence variation of target genes (*invA*, *ttrRSBCA*, *fimA*, *phoP*, *spvC*, and *agfA*) among these isolates; (2) predict the protein structures of the selected genes to identify the potential impact of the mutations on protein functions; and (3) provide a detailed comparative genomic analysis of these major serotypes. The study adds additional WGS data and new protein structure information to the *Salmonella* online database and provide insights for the improvement in the design of rapid molecular detection methods for *Salmonella*.

Results

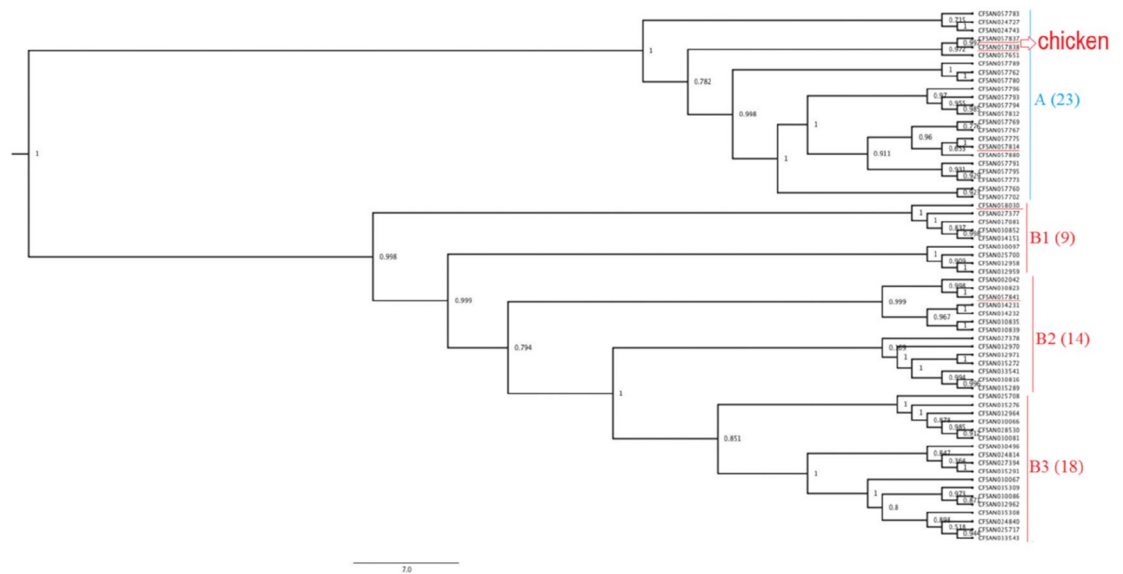
WGS and Phylogenetic analyses of the three *Salmonella* serotypes. The overall results of WGS, such as the number of assembled bases and N50 contig sizes, are summarized in Supplementary Tables 1–3 and Supplementary Fig. 1. In general, *Salmonella* genomes averaged at the size of approximately 5 Mb, the number of contigs ranged from 28 to 473, and the average depth of coverage ranged from 25× to 433× (Supplementary Tables 1–3). The WGS data of all the isolates studied here can be accessed from the NCBI SRA (<https://www.ncbi.nlm.nih.gov/sra>) with their accession numbers listed in Supplementary Tables 1–3.

Next, we constructed phylogenetic trees grouped by serotype to investigate the isolates' genetic diversity. The phylogenetic tree constructed for the 64 *Salmonella* ser. Enteritidis isolates was separated into two clades (clade A and clade B) with 552–565 SNPs (Fig. 1). Forty-one of the isolates were placed into clade B with 3 subclades, namely, clade B1 (9 isolates), clade B2 (14 isolates), and clade B3 (18 isolates), with 160–219 SNPs. The 40 *Salmonella* ser. Typhimurium isolates were grouped into two clades (clade A and clade B); each clade had three subclades (clade A1, A2, A3, and clade B1, B2, B3) (Fig. 2). Within the three subclades of clade A, we identified 48–81 SNPs; within the three subclades of clade B, we identified 673–1141 SNPs. All eight *Salmonella* ser. Typhimurium isolates from egg sources were placed into clade B2. The 39 *Salmonella* ser. Heidelberg isolates were grouped in three different clades, namely, clade A, clade B, and clade C (subclade A, B1, B2, C1, C2, and C3), with less than 100 SNPs (19–95 bp) among them (Fig. 3). Five egg-sourced isolates formed clade A. Eighteen isolates belonged to clade B, and 16 belonged to clade C. Clade B2 encompassed all nine chicken-sourced and two egg-sourced (CFSAN015479 and CFSAN033547) *Salmonella* ser. Heidelberg isolates; compared to other clades, this clade also had relatively high pairwise SNP distances from other clades (85–95 bp).

Distribution of selected target genes among three *Salmonella* serotypes. After sequencing and assembling the genomes, we performed BLAST analyses to investigate the existence of the selected target genes among *Salmonella* ser. Enteritidis, *Salmonella* ser. Typhimurium and *Salmonella* ser. Heidelberg isolates. The BLAST analyses showed that all 143 isolates contained the genes *invA*, *ttrRSBCA*, *phoP*, *fimA*, and *agfA* (data not shown). By contrast, not all the isolates carried the *spvC* gene. Specifically, this gene was present in 59/64 *Salmonella* ser. Enteritidis isolates and in 14/40 *Salmonella* ser. Typhimurium isolates whereas none of the 39 *Salmonella* ser. Heidelberg isolates carried *spvC*. We also identified isolates that carried only partial target gene sequences. For example, two *Salmonella* ser. Typhimurium isolates (CFSAN034209, CFSAN036243) carried a partial sequence of *fimA*, and six *Salmonella* ser. Heidelberg isolates carried a partial sequence of *ttrA* (CFSAN015377, CFSAN015378, CFSAN015380, CFSAN017093, CFSAN017094, CFSAN017095).

Mutations of the target genes. We identified numerous non-synonymous mutations and synonymous mutations among the selected target genes (Table 1). The most commonly detected non-synonymous mutations were changes between T and C (15 times), A and G (13 times), and T and G (8 times); the less frequent changes were those between G and C (4 times), A and C (4 times), and A and T (once). The mutation rates for A/T and G/C were 53.33% and 46.67%, respectively.

Regarding the mutations in each gene, we identified 13 unique non-synonymous mutations and 80 synonymous mutations in *invA* across all our isolates (data not shown). Although the mutations *invA*-Mut.1 and *invA*-Mut.2 were changes of C → G (nt position 770) and G → C (nt position 771), respectively, they resulted in the same amino acid (AA) change in *invA* protein, from threonine (T) to serine (S), at site 257 (Table 1).



Pairwise SNP distances (*Salmonella* ser. Enteritidis)

Clades	A	B1	B2	B3
A				
B1	552 (17)			
B2	562 (20)	215 (6)		
B3	565 (22)	219 (8)	160 (7)	

Figure 1. Phylogenetic tree of 64 *Salmonella* ser. Enteritidis isolates. Using core SNPs determined by FDA CFSAN SNP pipeline, the tree was constructed using the GTR-CAI model of FASTree 2. The analysis involved 552–565 total SNP positions. Clades and subclades are indicated in colors (A, blue; and B, red) with numbers of total isolates in each clade or subclade in parentheses. Isolates sourced from chicken samples were highlighted under red line, the rest were from egg samples.

In the *ttrS*, *ttrB*, *ttrC*, and *ttrA* genes we found 7, 2, 2, and 7 non-synonymous mutations, respectively. Most non-synonymous mutations in *ttrS* occurred in the *Salmonella* ser. Enteritidis and *Salmonella* ser. Heidelberg isolates. We did not identify non-synonymous mutations in the *ttrB* and *ttrA* genes among the *Salmonella* ser. Typhimurium isolates. Six *Salmonella* ser. Typhimurium isolates had a unique non-synonymous mutation in the *ttrS* gene (*ttrS*-Mut.4), and 12 *Salmonella* ser. Typhimurium isolates had a non-synonymous mutation in the *ttrC* gene (*ttrC*-Mut.1). We identified four *Salmonella* ser. Enteritidis isolates that carried mutations in *ttrB* (*ttrB*-Mut.1: CFSAN057880; *ttrB*-Mut.2: CFSAN057841, CFSAN030823, CFSAN002042). Both *ttrS*-Mut.1 (nt position 986, G → A) and *ttrS*-Mut.2 (nt position 987, T → C) resulted in an AA change of serine (S) → asparagine (N). No mutations were found in *ttrR* among all the isolates studied.

The *fimA* gene was found to have 13 non-synonymous mutations although most of them were identified in the *Salmonella* ser. Enteritidis and *Salmonella* ser. Heidelberg isolates; only two synonymous mutations separately found in one *Salmonella* ser. Enteritidis isolate (CFSAN030097, raw egg yolks) and in all *Salmonella* ser. Heidelberg isolates (data not shown). Notably, all *Salmonella* ser. Enteritidis and *Salmonella* ser. Heidelberg isolates carried one common nonsense mutation (*fimA*-Mut.6), C → T (nt position 496), resulting in a change at AA 166, namely, glutamine (Q) → stop codon (TAG).

Finally, the *phoP* gene exhibited one non-synonymous mutation across 25 *Salmonella* ser. Enteritidis isolates and four synonymous mutations among all *Salmonella* ser. Enteritidis isolates; no *phoP* mutation was found among the isolates in the other two serotypes. For the *agfA* and *spvC* genes, only six synonymous mutations occurred in the *Salmonella* ser. Heidelberg and *Salmonella* ser. Typhimurium isolates, and one synonymous mutation of the *spvC* gene occurred in *Salmonella* ser. Enteritidis isolates (data not shown). No non-synonymous mutations were observed for the *agfA* and *spvC* genes among the isolates studied.

Phylogenetic analyses for each target gene. Phylogenetic trees were then constructed for each of the selected target genes based on their nucleotide sequences (including synonymous/non-synonymous mutations) in the *Salmonella* isolates studied to investigate their genetic differences. We found disparities among all selected target genes (Figs. 4, 5, 6) across all three serotypes except *ttrR* gene. Four *Salmonella* ser. Enteritidis isolates (CFSAN030097, CFSAN025700, CFSAN032958, CFSAN032959) were different from the others based on the phylogenetic tree of the *fimA* gene, and the *Salmonella* ser. Typhimurium isolates were separated into three clusters (Fig. 4C). In the *phoP*-based phylogenetic tree, the isolates of *Salmonella* ser. Typhimurium and *Salmonella* ser. Heidelberg were in the same clade, and *Salmonella* ser. Enteritidis isolates were divided into two different clades (Fig. 5A). Variations were observed for the phylogenetic trees constructed for *ttrS* (Fig. 6A), *ttrB*

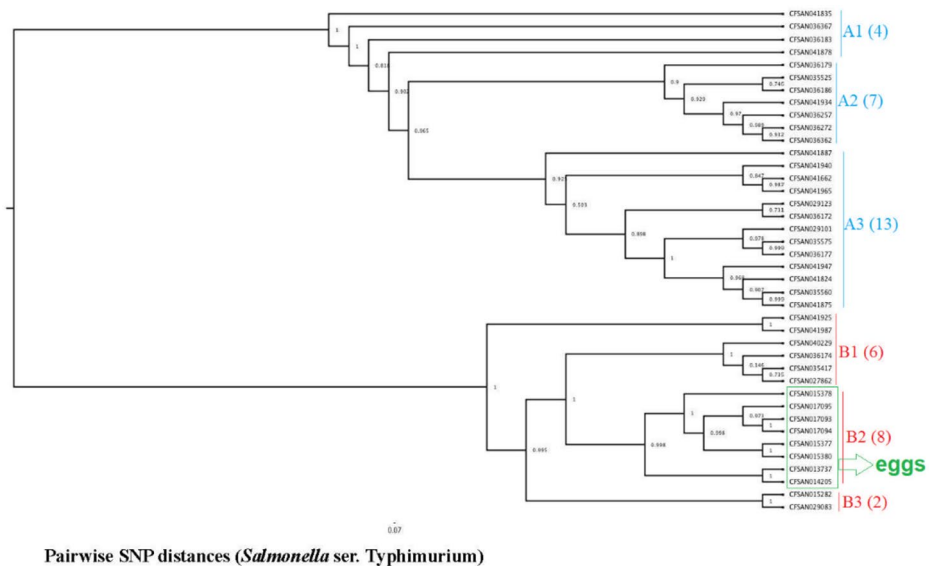


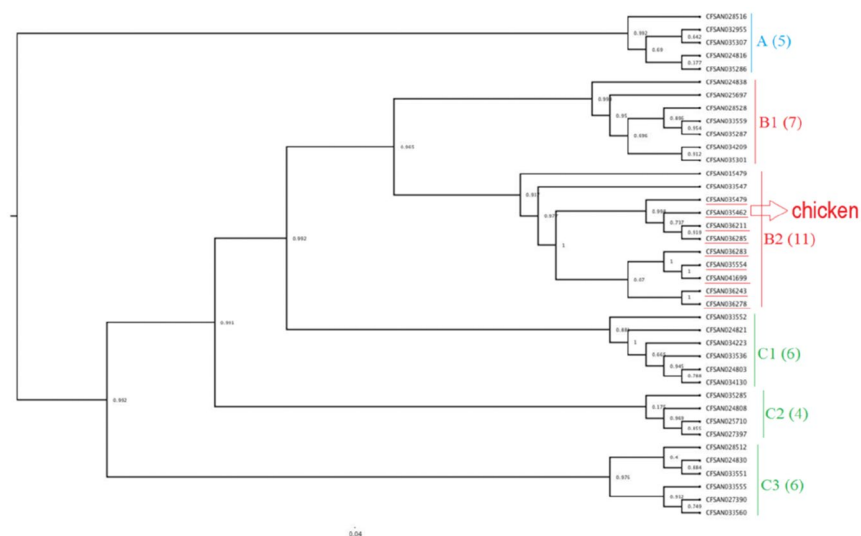
Figure 2. Phylogenetic tree of 40 *Salmonella* ser. Typhimurium isolates. Using core SNPs determined by FDA CFSAN SNP pipeline, the tree was constructed under the GTR-CAI model of FASTree 2. The analysis involved 673–1141 total SNP positions. Clades and subclades are indicated in colors (A, blue; and B, red) with numbers of total isolates in each clade or subclade in parentheses. Isolates sourced from eggs and clustered in subclade B2 are highlighted in green, the rest were from chicken samples.

(Fig. 6B), *ttrC* (Fig. 6C) and *ttrA* (Fig. 6D). As our set of isolates did not exhibit mutations in *ttrR*, we could not generate a tree based on that target gene.

Protein structure prediction. The protein structures for the genes *phoP* (Fig. 7C), *spvC* (Fig. 7D), *ttrS* (Fig. 8A), *ttrB* (Fig. 8B), *ttrC* (Fig. 8C), and *ttrA* (Fig. 8D) were successfully predicted with high confidence (>90%). The predicted protein structure of the *invA* gene showed a good Global Model Quality Estimation (GMQE) score of 0.43 and Qualitative Model Energy Analysis (QMEAN) score of 1.01, which covered 7 of the AA mutations revealed (Fig. 7A). Although, we tried to predict the protein structure resulting from the *fimA* gene using Phyre² and SWISS-MODEL, both software programs failed to generate an acceptable result. The predicted protein structure for the *fimA* gene, as generated by SWISS-MODEL, had a GMQE score of 0.99, but the low QMEAN score (-5.44) suggested that it may not be the best representation possible (Fig. 7B). However, the structure generated by Phyre² was worse and exhibited only less than 21% confidence (model not shown). Similarly, we did not finalize models for the protein structures of the *agfA* and *ttrR* genes, as the confidence and quality scores from both Phyre² and SWISS-MODEL were too low to proceed. Additionally, from the predicted structure (Figs. 7 and 8), evidently the two common secondary structures α -helix and β -sheet existed in each gene: *invA* (8 α -helices and 5 β -sheets), *fimA* (8 α -helices and 2 β -sheets), *phoP* (8 α -helices and 4 β -sheets), *spvC* (6 α -helices and 3 β -sheets), *ttrS* (14 α -helices and 6 β -sheets), *ttrB* (4 α -helices and 3 β -sheets), *ttrC* (12 α -helices and 0 β -sheets), and *ttrA* (15 α -helices and 12 β -sheets).

Discussion

Phylogenetic analyses of three *Salmonella* serotypes. WGS is the most powerful tool for bacterial genomic variation analyses because it is based on the complete bacterial genome at DNA level. WGS SNP analysis was used to ascertain that a new lineage of *Salmonella* ser. Enteritidis occurred and spread in Brazil after 1994 by investigating 256 *Salmonella* ser. Enteritidis isolates obtained over 48 years²³. In Denmark, reanalysis of isolates in eight previously reported outbreaks by WGS successfully discriminated 372 isolates of *Salmonella* ser. Typhimurium and its monophasic variants²⁴. Allard et al. (2013) sequenced and analysed 106 *Salmonella* ser. Enteritidis isolates from clinical, food and farm environments related to the 2010 shell egg outbreak in the U.S.; these isolates were classified into nine lineages and had a range of 100 to 600 diverse SNPs²⁵. Across the 143 genomes of *Salmonella* Enteritidis, Typhimurium and Heidelberg analyzed in the current report, *Salmonella*



Pairwise SNP distances (*Salmonella* ser. Heidelberg)

Clades	A	B1	B2	C1	C2	C3
A	43 (6)					
B1	89 (6)	85 (5)				
B2	39 (5)	45 (6)	90 (6)			
C1	26 (3)	46 (5)	91 (5)	41 (5)		
C2	19 (2)	49 (5)	95 (5)	44 (5)	32 (3)	
C3						

Figure 3. Phylogenetic tree of 39 *Salmonella* ser. Heidelberg isolates. Using core SNPs determined by FDA CFSAN SNP pipeline, the tree was constructed under the GTR-CAI model of FASTree 2. The analysis involved 19–95 total SNP positions. Clades and subclades are indicated in colors (A, blue; B, red; and C, green) with numbers of total isolates in each clade or subclade in parentheses. Isolates sourced from chicken and clustered in subclade B2 are highlighted in red, the rest were from egg samples.

ser. Typhimurium displayed the highest degree of genetic diversity, with a distance of 48–1152 SNPs among the different clades, while the value for *Salmonella* ser. Enteritidis was 160–565 SNPs, and that for *Salmonella* ser. Heidelberg was 19–95 SNPs (Figs. 1, 2, 3). Our results agree with a previous study in which *Salmonella* ser. Typhimurium had the highest phylogenetic diversity compared with *Salmonella* ser. Newport and *Salmonella* ser. Dublin²⁶. *Salmonella* ser. Typhimurium and its monophasic variant have quite distinctive evolutive pathways (exploration–exploitation pathways) in comparison with *Salmonella* ser. Enteritidis and *Salmonella* ser. Heidelberg, which could be one of the explanations for this phenomenon²⁷. Interestingly, in this study, all eight *Salmonella* ser. Typhimurium isolates from egg sources were placed into the same subclade (clade B2), whereas five egg sourced *Salmonella* ser. Heidelberg grouped into one clade (clade A) and nine chicken sourced isolates and two egg sourced isolates placed into the same subclade (clade B2). Unfortunately, there was limited epidemiological information available for these isolates to explore the relationships among these isolates.

DNA sequence variation of target genes among *Salmonella* isolates and their phylogenetic relationships.

The selected target genes/gene clusters *invA*, *ttrRSBCA*, *phoP*, *fimA*, *agfA*, and *spvC* are all virulence factors of *Salmonella*. The *spvC* gene is on the *Salmonella* virulence plasmid, while the others are located on chromosomes²⁸. Although the *spvABCD* gene cluster was reported to be highly conserved in *Salmonella* ser. Typhimurium, *Salmonella* ser. Dublin and *Salmonella* ser. Choleraesuis²⁹, our analysis revealed that the *spvC* gene was primarily found among *Salmonella* ser. Enteritidis isolates (92.19%), followed by *Salmonella* ser. Typhimurium isolates (35%). None of our *Salmonella* ser. Heidelberg isolates carried *spvC*, which is consistent with previous reports^{30–33}. Study had shown that *spvC* is essential for the virulence of *Salmonella* ser. Typhimurium in mice; the potential impact of the presence and absence of the *spvC* gene on the isolates' ability of causing infection warrants further investigation⁵⁵. The *ttrR* gene was the most highly conserved compared to the other *ttr* genes, as no mutations were found in *ttrR* among all the isolates studied. In addition, only selected alleles which were more frequently used in detection technology design of the selected target genes/gene clusters were investigated in the current project.

Analysis of the phylogenetic diversity of the serotypes based on the phylogenetic trees constructed using individual genes, showed that the serotypes differ in the genes *ttrS*, *ttrB*, *ttrC*, and *ttrA* (Fig. 6). Variations were observed in *Salmonella* ser. Enteritidis and *Salmonella* ser. Typhimurium, but the genes in *Salmonella* ser. Heidelberg were more stable. The *ttrR* gene was the most highly conserved compared to the other *ttr* genes, because no mutations were found in *ttrR* among all the isolates studied. As a gene cluster within the *Salmonella* pathogenicity

Genes	Non-synonymous mutation	Nucleotide (position)	NT change	Protein (position)	AA change	Associated isolates
<i>invA</i>	<i>invA</i> -Mut. 1	770	C → G	257	T → S	All isolates
	<i>invA</i> -Mut. 2	771	G → C	257	T → S	All isolates
	<i>invA</i> -Mut. 3	898	G → T	300	A → S	All isolates
	<i>invA</i> -Mut. 4	1012	A → T	338	T → S	All isolates
	<i>invA</i> -Mut. 5	1125	T → G	375	D → E	All isolates
	<i>invA</i> -Mut. 6	1198	G → C	400	V → L	All isolates
	<i>invA</i> -Mut. 7	1201	A → G	401	S → D	All isolates
	<i>invA</i> -Mut. 8	1202	G → A	401	S → D	All isolates
	<i>invA</i> -Mut. 9	1301	C → T	434	A → V	All isolates
	<i>invA</i> -Mut. 10	1302	T → G	434	A → V	All <i>Salmonella</i> ser. Enteritidis and <i>Salmonella</i> ser. Typhimurium isolates
	<i>invA</i> -Mut. 11	1325	C → T	442	T → I	All isolates
	<i>invA</i> -Mut. 12	1327	T → C	443	Y → H	All <i>Salmonella</i> ser. Enteritidis and <i>Salmonella</i> ser. Typhimurium isolates
	<i>invA</i> -Mut. 13	1332	T → G	444	H → Q	All isolates
<i>ttrS</i>	<i>ttrS</i> -Mut. 1	986	G → A	329	S → N	All <i>Salmonella</i> ser. Enteritidis isolates
	<i>ttrS</i> -Mut. 2	987	C → T	329	S → N	All <i>Salmonella</i> ser. Enteritidis isolates
	<i>ttrS</i> -Mut. 3	1117	T → G	373	C → G	All <i>Salmonella</i> ser. Enteritidis isolates
	<i>ttrS</i> -Mut. 4	1450	G → T	484	A → S	<i>Salmonella</i> ser. Enteritidis (CFSAN034231, CFSAN034232); <i>Salmonella</i> ser. Typhimurium (CFSAN017095, CFSAN017094, CFSAN017093, CFSAN015380, CFSAN015378, CFSAN015377)
	<i>ttrS</i> -Mut. 5	1465	A → G	489	I → V	All <i>Salmonella</i> ser. Enteritidis and <i>Salmonella</i> ser. Heidelberg isolates
	<i>ttrS</i> -Mut. 6	1486	A → C	496	N → H	All <i>Salmonella</i> ser. Enteritidis and <i>Salmonella</i> ser. Heidelberg isolates
	<i>ttrS</i> -Mut. 7	1502	T → C	501	V → A	All <i>Salmonella</i> ser. Enteritidis and <i>Salmonella</i> ser. Heidelberg isolates
<i>ttrB</i>	<i>ttrB</i> -Mut.1	389	T → C	130	I → T	<i>Salmonella</i> ser. Enteritidis (CFSAN057880)
	<i>ttrB</i> -Mut.2	650	C → A	217	P → Q	<i>Salmonella</i> ser. Enteritidis (CFSAN057841, CFSAN030823, CFSAN002042)
<i>ttrC</i>	<i>ttrC</i> -Mut.1	560	G → A	187	R → H	12 <i>Salmonella</i> ser. Typhimurium isolates ^a
	<i>ttrC</i> -Mut.2	740	G → T	247	C → F	All <i>Salmonella</i> ser. Enteritidis and <i>Salmonella</i> ser. Heidelberg isolates
<i>ttrA</i>	<i>ttrA</i> -Mut.1	488	T → C	163	V → A	All <i>Salmonella</i> ser. Enteritidis isolates
	<i>ttrA</i> -Mut.2	634	C → T	212	R → C	41 <i>Salmonella</i> ser. Enteritidis isolates ^b
	<i>ttrA</i> -Mut.3	1024	C → G	342	Q → E	All <i>Salmonella</i> ser. Enteritidis isolates
	<i>ttrA</i> -Mut.4	2057	A → G	686	H → R	All <i>Salmonella</i> ser. Enteritidis isolates
	<i>ttrA</i> -Mut.5	2066	C → T	689	A → V	All <i>Salmonella</i> ser. Enteritidis isolates
	<i>ttrA</i> -Mut.6	2123	A → G	708	H → R	All <i>Salmonella</i> ser. Heidelberg isolates
	<i>ttrA</i> -Mut.7	2170	A → G	724	T → A	All <i>Salmonella</i> ser. Heidelberg isolates
<i>fimA</i>	<i>fimA</i> -Mut.1	19	T → C	7	S → P	All <i>Salmonella</i> ser. Heidelberg isolates
	<i>fimA</i> -Mut.2	86	C → A	29	P → H	24 <i>Salmonella</i> ser. Typhimurium isolates ^c
	<i>fimA</i> -Mut.3	127	A → G	43	I → V	All <i>Salmonella</i> ser. Heidelberg isolates
	<i>fimA</i> -Mut.4	257	A → G	86	D → G	<i>Salmonella</i> ser. Enteritidis (CFSAN030097, CFSAN025700, CFSAN032858, CFSAN032959)
	<i>fimA</i> -Mut.5	319	C → T	107	P → S	<i>Salmonella</i> ser. Typhimurium (CFSAN017095, CFSAN017094, CFSAN017093, CFSAN015380, CFSAN015378, CFSAN015377)
	<i>fimA</i> -Mut.6	496	C → T	166	Q → Stop Codon	All <i>Salmonella</i> ser. Enteritidis and <i>Salmonella</i> ser. Heidelberg isolates
	<i>fimA</i> -Mut.7	630	A → C	210	Q → H	All <i>Salmonella</i> ser. Enteritidis and <i>Salmonella</i> ser. Heidelberg isolates
	<i>fimA</i> -Mut.8	709	T → C	237	Y → H	All <i>Salmonella</i> ser. Enteritidis isolates
	<i>fimA</i> -Mut.9	745	T → C	249	Y → H	All <i>Salmonella</i> ser. Heidelberg isolates
	<i>fimA</i> -Mut.10	751	T → C	251	F → L	All <i>Salmonella</i> ser. Heidelberg isolates
	<i>fimA</i> -Mut.11	794	T → G	265	V → G	All <i>Salmonella</i> ser. Enteritidis and <i>Salmonella</i> ser. Heidelberg isolates
	<i>fimA</i> -Mut.12	851	G → A	284	R → H	All <i>Salmonella</i> ser. Enteritidis isolates
	<i>fimA</i> -Mut.13	857	A → G	286	H → R	All <i>Salmonella</i> ser. Heidelberg isolates
<i>phoP</i>	<i>phoP</i> -Mut.1	268	G → A	90	E → K	25 <i>Salmonella</i> ser. Enteritidis isolates ^d

Table 1. Non-synonymous mutations observed in each target gene from all studied *Salmonella* isolates. ^a12 *Salmonella* ser. Typhimurium isolates: CFSAN040229, CFSAN036174, CFSAN035417, CFSAN027862, CFSAN017095, CFSAN017094, CFSAN017093, CFSAN015380, CFSAN015378, CFSAN015377, CFSAN014205, CFSAN013737. ^b41 *Salmonella* ser. Enteritidis isolates: CFSAN057841, CFSAN035309, CFSAN035308, CFSAN035291, CFSAN035289, CFSAN035276, CFSAN035272, CFSAN034232, CFSAN034231, CFSAN034151, CFSAN033543, CFSAN033541, CFSAN032971, CFSAN032970, CFSAN032964, CFSAN032962, CFSAN032959, CFSAN032958, CFSAN030852, CFSAN030839, CFSAN030835, CFSAN030823, CFSAN030816, CFSAN030496, CFSAN030097, CFSAN030086, CFSAN030081, CFSAN030067, CFSAN030066, CFSAN028530, CFSAN027394, CFSAN027378, CFSAN027377, CFSAN025717, CFSAN025708, CFSAN025700, CFSAN024840, CFSAN024814, CFSAN017081, CFSAN002042, CFSAN058030. ^c24 *Salmonella* ser. Typhimurium isolates: CFSAN041965, CFSAN041947, CFSAN041940, CFSAN041934, CFSAN041887, CFSAN041878, CFSAN041875, CFSAN041835, CFSAN041824, CFSAN041662, CFSAN036367, CFSAN036362, CFSAN036272, CFSAN036257, CFSAN036186, CFSAN036183, CFSAN036179, CFSAN036177, CFSAN036172, CFSAN035575, CFSAN035560, CFSAN035525, CFSAN029123, CFSAN029101. ^d25 *Salmonella* ser. Enteritidis isolates: CFSAN035309, CFSAN035308, CFSAN035291, CFSAN035289, CFSAN035276, CFSAN035272, CFSAN033543, CFSAN033541, CFSAN032971, CFSAN032970, CFSAN032964, CFSAN032962, CFSAN030816, CFSAN030496, CFSAN030086, CFSAN030081, CFSAN030067, CFSAN030066, CFSAN028530, CFSAN027394, CFSAN027378, CFSAN025717, CFSAN025708, CFSAN024840, CFSAN024814.

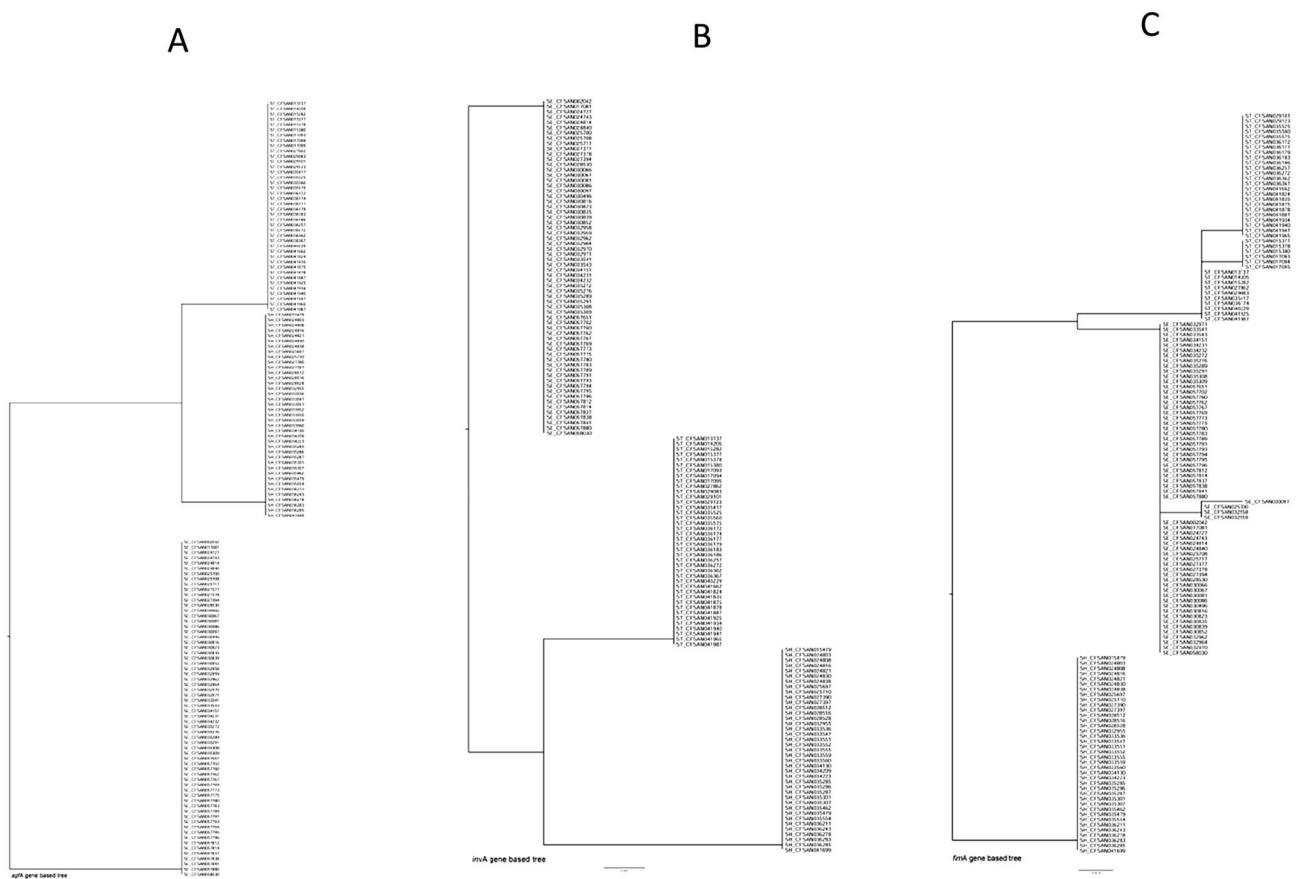


Figure 4. Phylogenetic trees of all sequenced *Salmonella* isolates based on nucleotide sequences of the target genes *agfA* (A), *invA* (B), and *fimA* (C). The trees were constructed by the neighbor joining method using CLC Genomics Workbench. The scale indicates the sequence percentage of base distance (percent divergence).

island (SPI, part of the flexible gene pool) on the chromosome, the *ttrRSBCA* locus has been shown to be transferable through horizontal gene transfer (HGT) events such as transfer by phages or conjugative transposons. This might be one of the reasons that many PCR protocols have an extra target (such as *invA*) in combination with the *ttrRSBCA* gene to reinforce the specificity of PCR detection^{9,34}. In addition, the presence of mutations in these genes could also be an underlying reason.

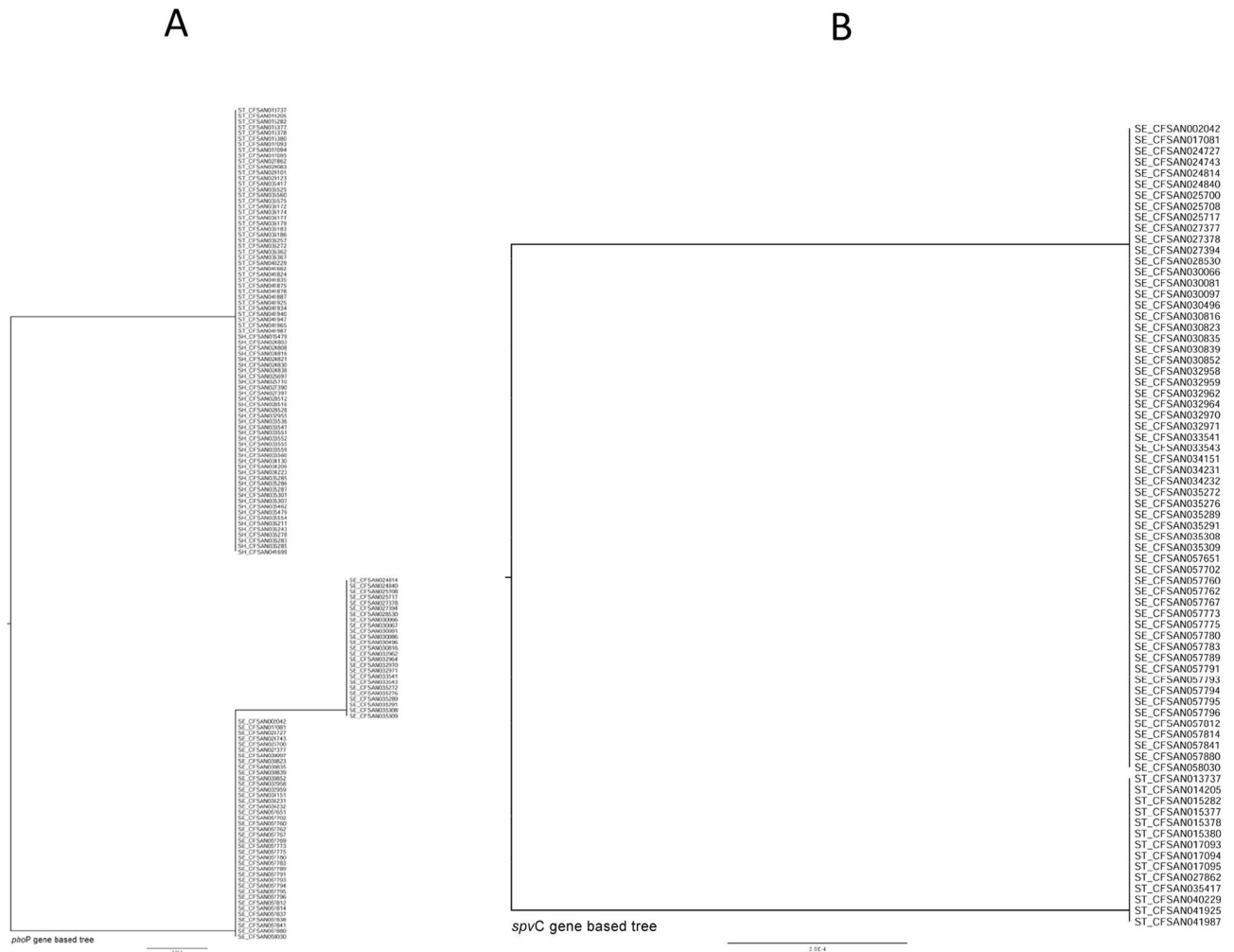


Figure 5. Phylogenetic trees of all sequenced *Salmonella* isolates based on the nucleotide sequences of the target genes *phoP* (A) and *spvC* (B). The trees were constructed by the neighbor joining method using CLC Genomics Workbench. The scale indicates the sequence percentage of base distance (percent divergence).

Finally, our study also demonstrated the presence of partial *fimA* (two isolates) and *ttrA* (six isolates) gene sequences among some *Salmonella* isolates and all studied *Salmonella* isolates contained the genes *invA*, *phoP*, and *agfA*. Phylogenetic trees generated based on both the *invA* and *agfA* genes clearly comprised three lineages of three serotypes (Fig. 4A and B). Interestingly, the *phoP*-based phylogenetic tree divided the *Salmonella* ser. Enteritidis isolates into two different clades (Fig. 5A). In this case, along with the analysis of the *fimA*, *spvC*, and *ttrR* genes (Figs. 4C, 5B, 6), we revealed that the *invA*, *phoP*, and *agfA* genes have higher discriminatory power to diagnose *Salmonella* at the genus level than other selected genes.

Mutations of target genes and prediction of protein structures. Among the 143 *Salmonella* isolates investigated in this study, both synonymous and non-synonymous mutations in the selected target genes were discovered. Among the total 45 non-synonymous mutations observed (Table 1), the top two mutations occurred with an AA change between the T and C alleles (15 times) and the A and G alleles (13 times). The substitution rates of A/T and G/C were approximately equal. Similarly, in a study involving 106 *Salmonella* ser. Enteritidis isolates, 55 non-synonymous mutations were discovered, and the top two mutations also occurred between the alleles T ↔ C (27 times) and A ↔ G (15 times)²⁵. These results are in accordance with the biased gene conversion model, in which AT → GC mutations have a higher probability of being transmitted to the next generation, as an AT/GC heterozygote produces more gametes carrying G or C than those carrying A or T, presumably through the GC-biased repair of A:C and G:T mismatches in heteroduplexed recombination intermediates^{35,36}. In addition, polymorphisms in an organism result from mutation, selection, and other processes such as biased gene conversion, which favors the transmission of G/C over A/T alleles^{35,36}.

In the successfully predicted protein structures of the target genes (Figs. 7 and 8), the overwhelming majority of the mutations were located in the main domain of the structure, except for *fimA*-Mut.1, which occurred close to the C-terminal, and *fimA*-Mut.13, which occurred close to the N-terminal. The mutations mainly occurred in the α-helix and β-sheet which are key elements for maintaining the conformation of proteins, therefore, such mutations could interfere with the hydrogen bonding between main-chain amide and carbonyl groups and their

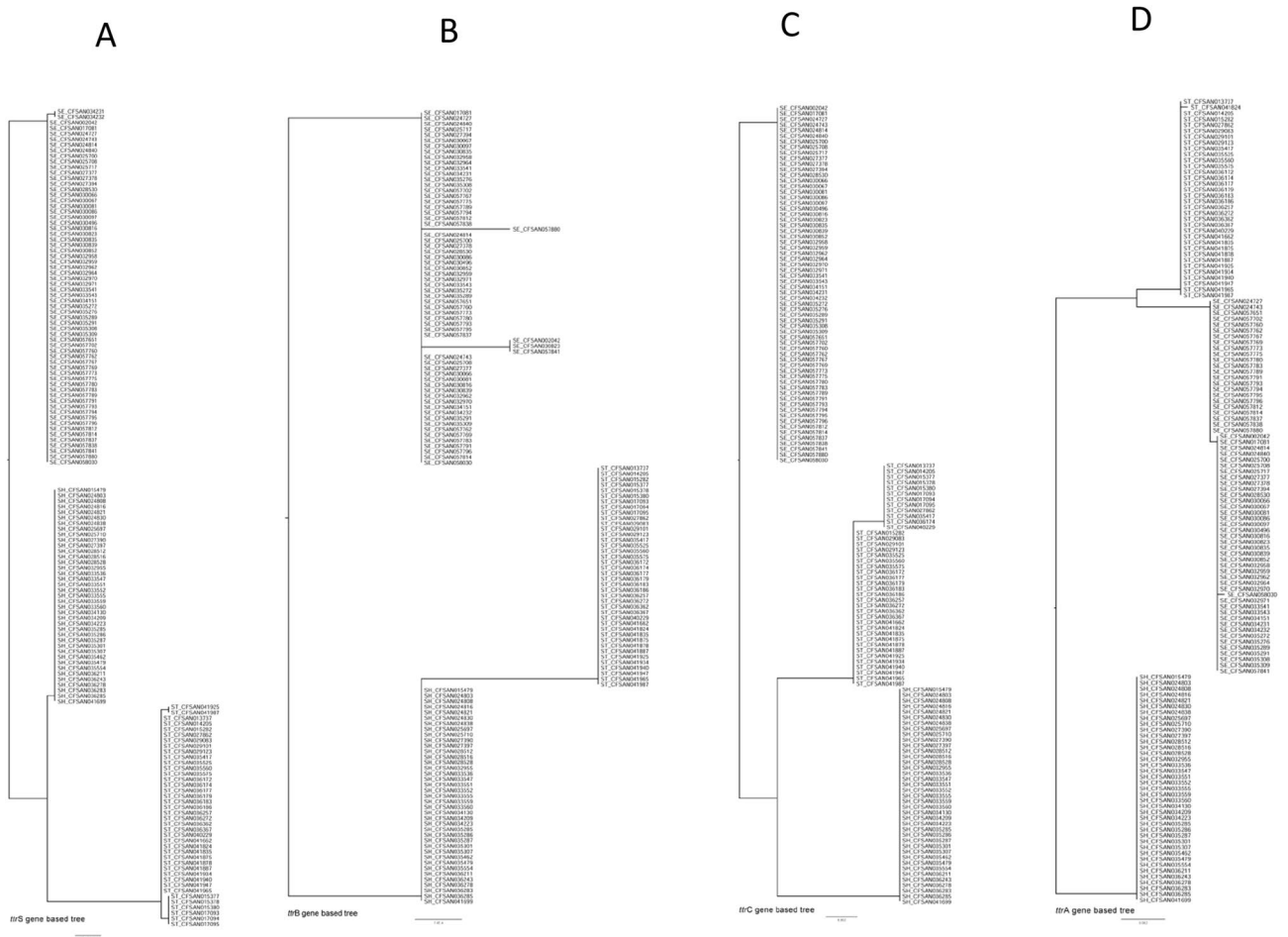


Figure 6. Phylogenetic trees of all sequenced *Salmonella* isolates based on the nucleotide sequences of the target genes *ttrS* (A), *ttrB* (B), *ttrC* (C), and *ttrA* (D). The trees were constructed by the neighbor joining method using CLC Genomics Workbench. The scale indicates the sequence percentage of base distance (percent divergence).

corresponding representations. It is well known that C-terminal sequence is an important structural and functional site of proteins and peptides whereas N-terminal influences the overall biological function of the protein.

The *invA* gene encodes an N-terminal integral membrane domain and a C-terminal cytoplasmic domain that is proposed to form part of a docking platform. As *invA* is essential for *Salmonella* to gain access to epithelial cells, isolates with non-synonymous *invA* mutant may have reduced virulence^{37,38}. Since *invA* gene has been well acknowledged as an effective target to detect *Salmonella*, it seems that the gene mutations happened to *invA* have less impact on the protein functions. It will take further experiments and data to prove this theory and reveal the mechanisms.

In terms of the number of mutations present in *ttrRSBCA*, we speculated that the *ttrR* gene was the most highly conserved, followed by the *ttrB* and *ttrC* genes, which have been successfully used in real-time PCR detection of *Salmonella* in food^{10,34}. Differences in the prevalence of mutations in the *ttrRSBCA* cluster may be related to the distinct functions and loci of these genes. The *ttrR* and *ttrS* genes are components of the *ttrSR* two-component regulatory system for functional tetrathionate reductase expression, while *ttrA*, *ttrB* and *ttrC* are tetrathionate reductase structural genes. The *ttrB* iron-sulphur clusters probably function in the transfer of electrons from *ttrC* to *ttrA*³⁹.

Among the 64 *Salmonella* ser. Enteritidis isolates studied, 25 had one mutation each in the *phoP* gene (G → A, nt 268). Allard et al. (2013) also observed the same non-synonymous mutation in the *phoP* gene in *Salmonella* ser. Enteritidis isolates from egg-associated samples²⁵. The *phoP* protein has a conserved N-terminal domain with an essential aspartate residue and a C-terminal domain that binds DNA. The *phoP/Q* two-component system, encoded by *phoP* and *phoQ*, controls more than 40 genes, such as *prgs*, *pagO*, *pagC*, and *pagD*, which regulate the host inflammatory response, lipopolysaccharide (LPS) formation, and extracellular protein transport and promote virulence and intracellular survival^{40–43}. This system may also play a specific role in *Salmonella* ser. Enteritidis pathogenicity in mice⁴⁴. A *phoP*-based LAMP assay has been developed to effectively detect *Salmonella* in food samples¹³. The stable peculiarity (one mutation has been found) of *phoP* gene also demonstrated the potential of *phoP* gene as target gene for detecting *Salmonella*.

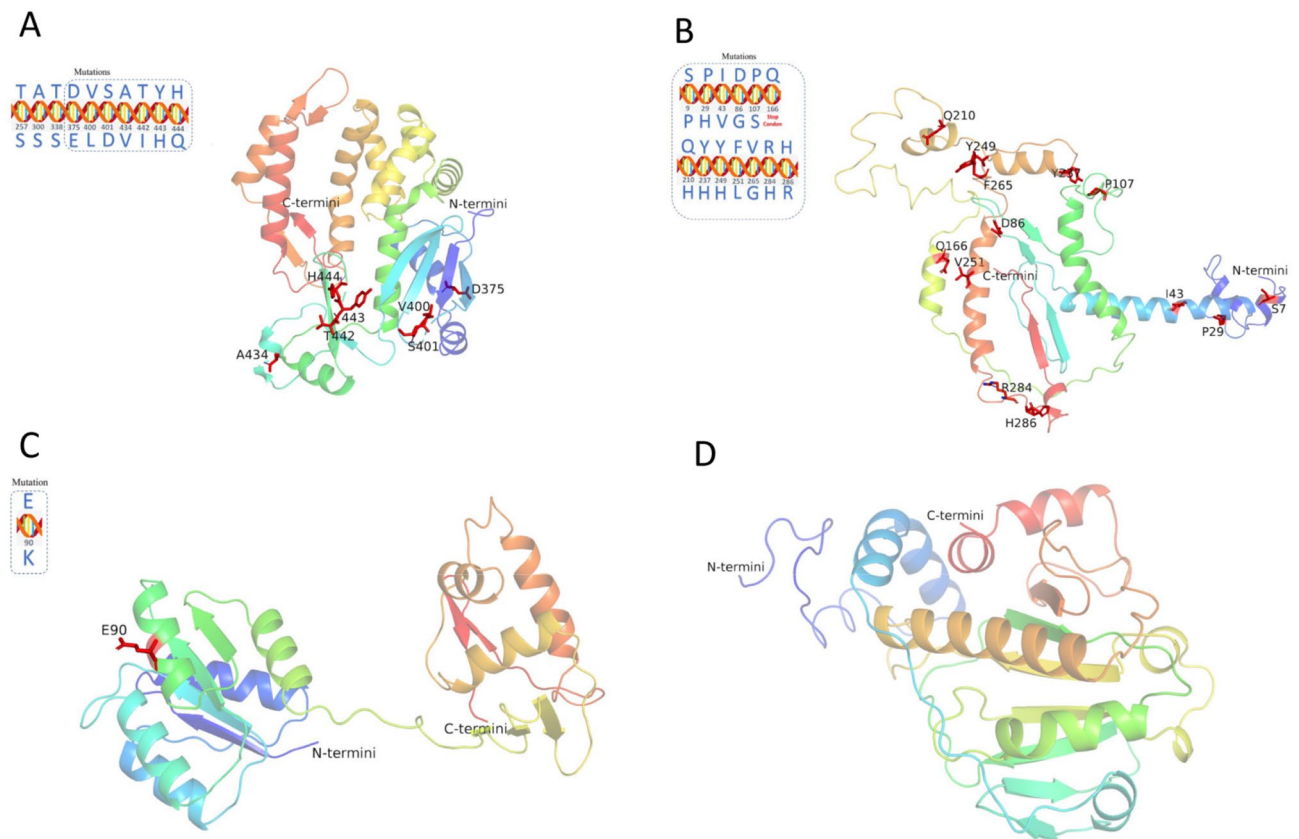


Figure 7. Predicted protein structures of the target genes *invA* (A), *fimA* (B), *phoP* (C), and *spvC* (D). Positions of amino acid mutations in each target protein are shown in the circled box and labeled in red in the predicted model.

It was reported that the *fimA* gene contains sequences unique to *Salmonella* strains and is an effective target for detecting *Salmonella* in feed and food samples^{16,45}. However, in this study, we not only found the *fimA* mutations happened close to N/C-terminal, but all *Salmonella* ser. Enteritidis and *Salmonella* ser. Heidelberg isolates carried one common nonsense mutation (*fimA*-Mut.6, Q → stop codon), which indicated that the *fimA* gene has questionable sites for being used as a target of method design to detect *Salmonella*. The *fimA* gene, encoding a major fimbria unit, was mapped within the *fim* gene cluster for the chaperone–usher pathway for the assembly and secretion of multi-subunit appendages (type I pili/fimbriae). The type I pili consists of a helical rod-like structure (*fimA* and *papA*) and a flexible tip that contains the minor pilus subunits (*fimE*, *fimG* and *fimH*, *papE*, *papF*, *papG*, *papK*). Solved crystal structures have shown the elongation complex *fimD*–*fimH*–*fimG*–*fimF*–*fimC* and the next subunit–chaperone complex of *fimA*–*fimC* in the chaperone–usher pathway. The 3D structure of *fimA* modelled on a *fimH*-G1 template indicates that the interface between the subunits contains small hydrophobic or polar residues such as alanine (A), serine (S) and threonine (T)^{46–48}. This may provide an explanation for the revealed non-synonymous mutation at the N-terminus (S → P) of the predicted *fimA* structure.

No non-synonymous mutations were detected in the *agfA* and *spvC* genes. Our attempt to predict protein structure of *agfA* was unsuccessful. Limited reports on *agfA* gene call for further research on this gene. The *agfBCA* operon encodes thin aggregative fimbriae/curli (formerly SEF17), and the thin aggregative fimbriae are primarily comprised of *agfA* subunits^{49,50}. Although thin aggregative fimbriae are produced by most *Salmonella* and *Escherichia coli* isolates⁵¹ and a high thin aggregative fimbriae sequence similarity was found between *Salmonella* ser. Enteritidis SEF17 fimbriae and *E. coli* curli⁴⁹. Doran et al. (1993) reported that *agfA*-based nucleotide probes hybridized only to *Salmonella* DNA⁵². The *agfA* gene has been successfully used as a target for *Salmonella* detection⁹. The *spv* genes, including *spvA*, *spvB*, *spvC*, *spvD*, and *spvR*, are often carried on a large *Salmonella* virulence plasmid. However, in some serotypes, they are integrated into the chromosome⁵³. The *spvC* gene, as a *Salmonella* effector with phosphothreonine lyase activity towards host mitogen-activated protein kinases, can be secreted in vitro by the SPI-1 and SPI-2 type III secretion systems⁵⁴. This gene is essential for the full virulence of *Salmonella* ser. Typhimurium in mice⁵⁵. Oligonucleotide insertions in *spvC* were shown to be nonpolar⁵⁶.

Currently, limited information is known about the complex functions of the target genes frequently used for the detection of the above mentioned *Salmonella* isolates. Our study provided comparison of 10 target genes from the perspective of DNA sequence and protein structure. More efforts should be made to determine whether these mutations can further affect the protein functions. And further biological, molecular, and functionality research would help fill the knowledge gaps in this area. We found both non-synonymous and synonymous mutation rates vary among the target genes frequently used based on the three serotypes studied. Large scale investigation of

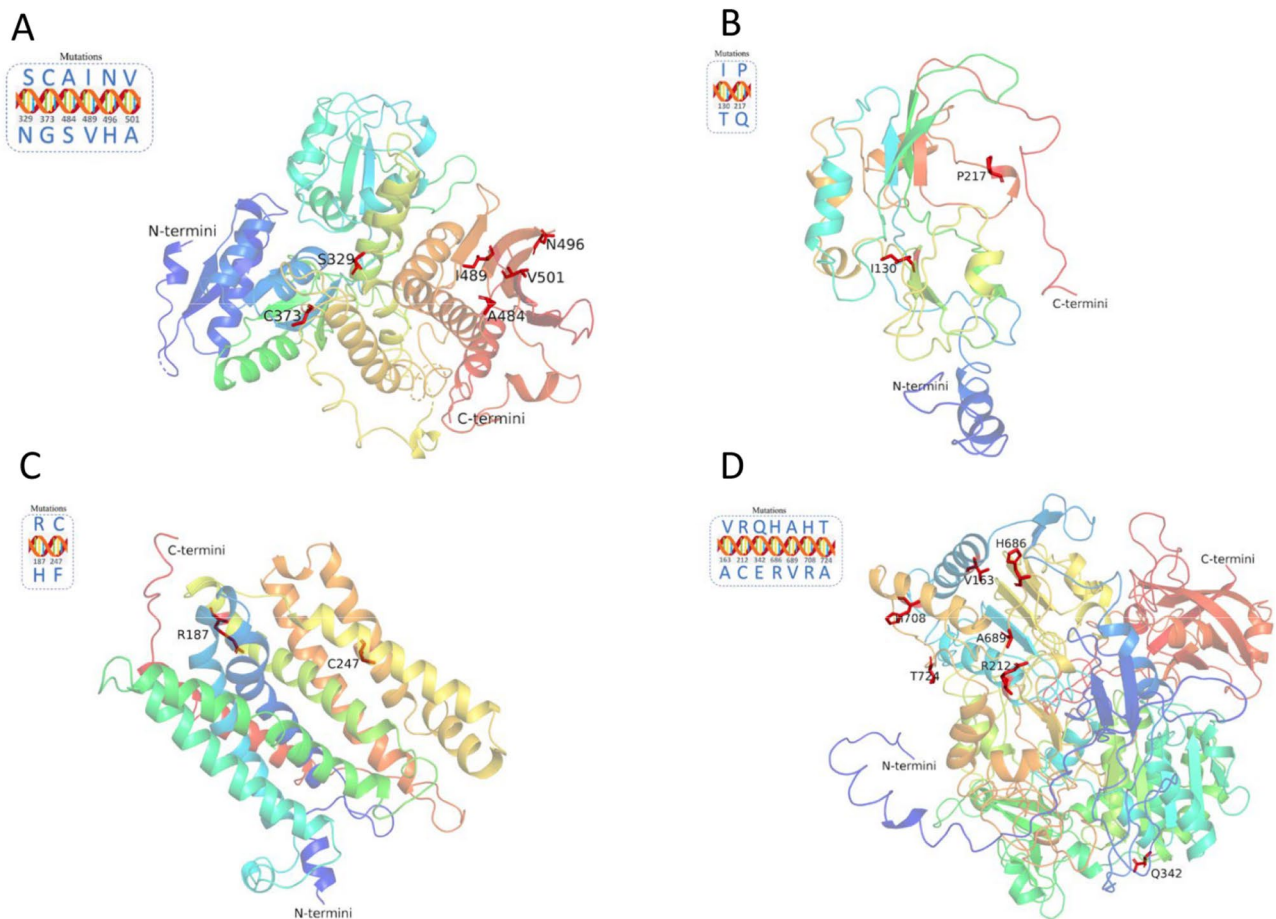


Figure 8. Predicted protein structures of the target genes *ttrS* (A), *ttrB* (B), *ttrC* (C), and *ttrA* (D). Positions of amino acid mutations in each target protein are shown in the circled box and labeled in red in the predicted model.

more serotypes regarding mutation rates are needed to determine which genes are more suitable for use as detection targets. With the use and sharing of this new information about these target genes in the future, the ability to identify and investigate *Salmonella* infections by comparing gene sequence data will be greatly enhanced.

Materials and methods

Salmonella isolates. We selected 143 *Salmonella* isolates from chicken/duck eggs (including raw egg whites, raw egg yolks, raw whole eggs, pecked eggs, egg slurry, egg salad, frozen liquid egg, cooked quail eggs, salted duck eggs, duck egg yolks, frozen salted duck yolks) and chicken (chicken, chicken jerky, chicken breast): 64 *Salmonella* ser. Enteritidis isolates were collected from 1995 to 2016 (Supplementary Table 1) as described previously⁵⁷; 40 *Salmonella* ser. Typhimurium isolates were collected from 2001 to 2010 (Supplementary Table 2); and 39 *Salmonella* ser. Heidelberg isolates were collected from 2003 to 2013 (Supplementary Table 3). The isolates were originally collected in the U.S., Brazil, China, and Chile. All isolates were from the historical collection of the Division of Microbiology (DM), Office of Regulatory Science (ORS), Center for Food Safety and Applied Nutrition (CFSAN), U.S. FDA. Serotypes of the isolates were reconfirmed using the Kauffmann and White classification scheme (https://www.pasteur.fr/sites/default/files/veng_0.pdf). All isolates were cultured overnight at 37 ± 2 °C in tryptic soy broth (Becton Dickinson, Franklin Lakes, NJ, USA) for DNA extraction.

Whole genome sequencing and creation of phylogenetic trees. Genomic DNA from each isolate was extracted using the DNeasy Blood and Tissue Kit (Qiagen, Valencia, CA) and quantified using a Qubit 3.0 fluorometer (Life Technologies, MD). Libraries were prepared according to Nextera XT protocols and then sequenced on the Illumina MiSeq/NextSeq 500 platform using the NextSeq 500/550 High Output Kit v2 following the manufacturer's instructions (Illumina, San Diego, CA). All genomes were sequenced with 300 cycles, pair-end library with coverage depth of $> 30 \times$ at FDA CFSAN. The Illumina reads were assembled de novo using CLC Genomics Workbench v9 (Qiagen Bioinformatics, Redwood City, CA). The Sequence Read Archive (SRA) Toolkit 2.8.1–3 (<https://trace.ncbi.nlm.nih.gov/Traces/sra/?view=software>) was used to download and convert SRR files to FASTQ files. Using the FDA CFSAN SNP pipeline⁵⁸, we acquired the SNP matrix and generated the phylogenetic tree under the GTR + CAT model of FASTtree2⁵⁹. The complete genomes of *Salmonella* ser. Enteritidis P125109 (NC_011294.1), *Salmonella* ser. Typhimurium LT2 (NC_003197.2), and *Salmonella* ser.

Heidelberg A3ES40 (NZ_CP016561.1) were utilized as the references for constructing the phylogenetic trees of *Salmonella* ser. Enteritidis, *Salmonella* ser. Typhimurium, and *Salmonella* ser. Heidelberg, respectively. FigTree software was used to visualize the generated phylogenetic trees (<http://tree.bio.ed.ac.uk/software/figtree/>). The WGS data of all the isolates studied here can be accessed from the NCBI SRA (<https://www.ncbi.nlm.nih.gov/sra>). Their accession numbers are listed (Supplementary Tables 1–3).

Identifying nucleotide mutations and predicting the protein structures of the selected genes.

The DNA sequences of the genes *invA*, *ttrRSBCA*, *phoP*, *fimA*, *agfA* and *spvC* from each isolate were aligned using MEGA v7 (<https://www.megasoftware.net/>) to the reference genes from NCBI: *invA* (GenBank: DQ644631.1), *ttrRSBCA* (GenBank: AF282268.1), *phoP* (GenBank: NC_003197.2, 1,318,679–1,319,365), *fimA* (GenBank: M18283.1), *agfA* (GenBank: U43280.1), and *spvC* (GenBank: HM044662.1). Protein structure prediction for *invA* and *fimA* was performed with SWISS-MODEL (Swiss Institute of Bioinformatics, Basel, Switzerland)⁶⁰, and that for other genes was performed using Phyre² (Structural Bioinformatics Group, Imperial College, London)⁶¹; all the structure illustrations were drawn using PyMOL v2.2 (<https://pymol.org/2/>).

Received: 17 July 2019; Accepted: 7 October 2021

Published online: 22 October 2021

References

- Centers for Disease Control and Prevention (CDC). *Burden of Foodborne Illness: Findings*, <https://www.cdc.gov/foodborneburden/2011-foodborne-estimates.html> (2016).
- World Health Organization (WHO). *WHO estimates of the global burden of foodborne diseases*, http://www.who.int/foodsafety/publications/foodborne_disease/fergreport/en/ (2015).
- Chiang, Y.-C. *et al.* Designing a biochip following multiplex polymerase chain reaction for the detection of *Salmonella* serovars Typhimurium, Enteritidis, Infantis, Hadar, and Virchow in poultry products. *J. Food Drug Anal.* **26**, 58–66 (2018).
- Gast, R. K. *et al.* Frequency and duration of fecal shedding of *Salmonella* serovars Heidelberg and Typhimurium by experimentally infected laying hens housed in enriched colony cages at different stocking densities. *Avian Dis.* **61**, 366–371 (2017).
- Chittick, P., Sulka, A., Tauxe, R. V. & Fry, A. M. A summary of national reports of foodborne outbreaks of *Salmonella* Heidelberg infections in the United States: clues for disease prevention. *J. Food Prot.* **69**, 1150–1153 (2006).
- Food and Agriculture Organization of the United Nations: Statistics Division (FAOSTAT). *Production: Livestock Primary: Eggs Primary*, <http://www.fao.org/faostat/en/#data/QL> (2018).
- International Egg commission (IEC). *Global egg production dynamics - past, present and future of a remarkable success story*, http://www.internationalegg.com/wp-content/uploads/2015/08/Economics-Report-StatsReportSept14_web.pdf (2014).
- Afroj, S. *et al.* Simultaneous detection of multiple *Salmonella* serovars from milk and chicken meat by real-time PCR using unique genomic target regions. *J. Food Prot.* **80**, 1944–1957 (2017).
- Crăciunaș, C., Keul, A.-L., Flonta, M. & Cristea, M. DNA-based diagnostic tests for *Salmonella* strains targeting *hilA*, *agfA*, *spvC* and *sef* genes. *J. Environ. Manag.* **95**, S15–S18 (2012).
- Dmitric, M. *et al.* In-house validation of real-time PCR methods for detecting the *invA* and *ttr* genes of *Salmonella* spp. in food. *J. Food Process. Preserv.* **42**, e13455 (2018).
- Josefsen, M. H., Krause, M., Hansen, F. & Hoorfar, J. Optimization of a 12-hour TaqMan PCR-based method for detection of *Salmonella* bacteria in meat. *Appl. Environ. Microbiol.* **73**, 3040–3048 (2007).
- Levin, R. E. The use of molecular methods for detecting and discriminating *Salmonella* associated with foods—a review. *Food Technol.* **23**, 313–367 (2009).
- Li, X. *et al.* A loop-mediated isothermal amplification method targets the *phoP* gene for the detection of *Salmonella* in food samples. *Int. Food Microbiol.* **133**, 252–258 (2009).
- Malorny, B., Bunge, C. & Helmuth, R. A real-time PCR for the detection of *Salmonella* Enteritidis in poultry meat and consumption eggs. *J. Microbiol. Methods.* **70**, 245–251 (2007).
- Malorny, B. *et al.* Diagnostic real-time PCR for detection of *Salmonella* in food. *Appl. Environ. Microbiol.* **70**, 7046–7052 (2004).
- Moreira, Â. N. *et al.* Detection of *Salmonella* Typhimurium in raw meats using in-house prepared monoclonal antibody coated magnetic beads and PCR assay of the *fimA* gene. *J. Immunoass. Immunoch.* **29**, 58–69 (2007).
- Reynisson, E., Josefsen, M. H., Krause, M. & Hoorfar, J. Evaluation of probe chemistries and platforms to improve the detection limit of real-time PCR. *J. Microbiol. Methods.* **66**, 206–216 (2006).
- Zahraei, S. T., Tadjbakhsh, H., Atashparvar, N., Nadalian, M. & Mahzounieh, M. Detection and identification of *Salmonella* Typhimurium in bovine diarrhoeic fecal samples by immunomagnetic separation and multiplex PCR assay. *Zoonoses Public Health* **54**, 231–236 (2007).
- U. S. Food and Drug Administration (FDA). *GenomeTrakr Fast Facts*, <https://www.fda.gov/Food/FoodScienceResearch/WholeGenomeSequencingProgramWGS/ucm403550.htm> (2018).
- National Center for Biotechnology Information (NCBI). *Pathogen Detection*, <https://www.ncbi.nlm.nih.gov/pathogens/> (2018).
- Inns, T. *et al.* Prospective use of whole genome sequencing (WGS) detected a multi-country outbreak of *Salmonella* Enteritidis. *Epidemiol. Infect.* **145**, 289–298 (2017).
- Ford, L. *et al.* Seven *Salmonella* Typhimurium outbreaks in Australia linked by trace-back and whole genome sequencing. *Foodborne Pathog. Dis.* **15**, 285–292 (2018).
- Campioni, F. *et al.* Changing of the genomic pattern of *Salmonella* Enteritidis strains isolated in Brazil over a 48 year-period revealed by Whole Genome SNP analyses. *Sci. Rep.* **8**, 10478 (2018).
- Gymoese, P. *et al.* Investigation of outbreaks of *Salmonella enterica* serovar Typhimurium and its monophasic variants using whole-genome sequencing, Denmark. *Emerg. Infect. Dis.* **23**, 1631 (2017).
- Allard, M. W. *et al.* On the evolutionary history, population genetics and diversity among isolates of *Salmonella* Enteritidis PFGE pattern JEGX01.0004. *PloS One* **8**, e55254 (2013).
- Carroll, L. M. *et al.* Whole-genome sequencing of drug-resistant *Salmonella enterica* isolated from dairy cattle and humans in New York and Washington states reveals source and geographic associations. *Appl. Environ. Microbiol.*, AEM. 00140-00117 (2017).
- Cliff, O. M. *et al.* Inferring evolutionary pathways and directed genotype networks of foodborne pathogens. *PLoS Comput. Biol.* **16**, e1008401. <https://doi.org/10.1371/journal.pcbi.1008401> (2020).
- The virulence Factor Database (VFDB). *Virulence factors of pathogenic bacteria*, <http://www.mgc.ac.cn/cgi-bin/VFs/genus.cgi?Genus=Salmonella> (2018).

29. Gulig, P. A. *et al.* Molecular analysis of spv virulence genes of the *Salmonella* virulence plasmids. *Mol. Microbiol.* **7**, 825–830 (1993).
30. Araque, M. Nontyphoid *Salmonella* gastroenteritis in pediatric patients from urban areas in the city of Mérida, Venezuela. *J. Infect. Dev. Ctries.* **3**, 028–034 (2009).
31. Chaudhary, J., Nayak, J., Brahmabhatt, M. & Makwana, P. Virulence genes detection of *Salmonella* serovars isolated from pork and slaughterhouse environment in Ahmedabad, Gujarat. *Vet. World* **8**, 121 (2015).
32. Castilla, K. S., Ferreira, C. S. A., Moreno, A. M., Nunes, I. A. & Ferreira, A. J. P. Distribution of virulence genes *sefC*, *pefA* and *spvC* in *Salmonella* Enteritidis phage type 4 strains isolated in Brazil. *Braz. J. Microbiol.* **37**, 135–139 (2006).
33. Swamy, S. C., Barnhart, H. M., Lee, M. D. & Dreesen, D. W. Virulence determinants *invA* and *spvC* in salmonellae isolated from poultry products, wastewater, and human sources. *Appl. Environ. Microbiol.* **62**, 3768–3771 (1996).
34. González-Escalona, N., Brown, E. W. & Zhang, G. Development and evaluation of a multiplex real-time PCR (qPCR) assay targeting *ttrRSBCA* locus and *invA* gene for accurate detection of *Salmonella* spp. in fresh produce and eggs. *Food Res. Int.* **48**, 202–208 (2012).
35. Galtier, N. & Duret, L. Adaptation or biased gene conversion? Extending the null hypothesis of molecular evolution. *Trends Genet.* **23**, 273–277 (2007).
36. Lynch, M. & Walsh, B. *The origins of genome architecture*. Vol. 98 (Sinauer Associates Sunderland (MA), 2007).
37. Galan, J. E. & Curtiss, R. Cloning and molecular characterization of genes whose products allow *Salmonella* Typhimurium to penetrate tissue culture cells. *Proc. Natl. Acad. Sci.* **86**, 6383–6387 (1989).
38. Worrall, L. J., Vuckovic, M. & Strynadka, N. C. Crystal structure of the C-terminal domain of the *Salmonella* type III secretion system export apparatus protein *InvA*. *Protein Sci.* **19**, 1091–1096 (2010).
39. Hensel, M., Hinsley, A. P., Nikolaus, T., Sawers, G. & Berks, B. C. The genetic basis of tetrathionate respiration in *Salmonella* Typhimurium. *Mol. Microbiol.* **32**, 275–287 (1999).
40. Behlau, I. & Miller, S. I. A *PhoP*-repressed gene promotes *Salmonella* Typhimurium invasion of epithelial cells. *J. Bacteriol.* **175**, 4475–4484 (1993).
41. Miller, S. I., Kukral, A. M. & Mekalanos, J. J. A two-component regulatory system (*phoP/phoQ*) controls *Salmonella* Typhimurium virulence. *Proc. Natl. Acad. Sci.* **86**, 5054–5058 (1989).
42. Pegues, D. A., Hantman, M. J., Behlau, I. & Miller, S. I. *PhoP/PhoQ* transcriptional repression of *Salmonella* Typhimurium invasion genes: evidence for a role in protein secretion. *Mol. Microbiol.* **17**, 169–181 (1995).
43. Tang, T., Cheng, A., Wang, M. & Li, X. Reviews in *Salmonella* Typhimurium *PhoP/PhoQ* two-component regulatory system. *Rev. Med. Microbiol.* **24**, 18–21 (2013).
44. Silva, C. A. *et al.* Infection of mice by *Salmonella enterica* serovar Enteritidis involves additional genes that are absent in the genome of serovar Typhimurium. *Infect. Immun.* **80**, 839–849 (2012).
45. Cohen, H. J., Mechanda, S. M. & Lin, W. PCR amplification of the *fimA* gene sequence of *Salmonella* Typhimurium, a specific method for detection of *Salmonella* spp. *Appl. Environ. Microbiol.* **62**, 4303–4308 (1996).
46. Costa, T. R. *et al.* Secretion systems in Gram-negative bacteria: structural and mechanistic insights. *Nat. Rev. Microbiol.* **13**, 343 (2015).
47. Hahn, E. *et al.* Exploring the 3D molecular architecture of *Escherichia coli* type 1 pili. *J. Mol. Biol.* **323**, 845–857 (2002).
48. Rossolini, G. M., Muscas, P., Chiesurin, A. & Satta, G. Analysis of the *Salmonella* *fim* gene cluster: identification of a new gene (*fimI*) encoding a fimbrin-like protein and located downstream from the *fimA* gene. *FEMS Microbiol. Lett.* **114**, 259–265 (1993).
49. Collinson, S. K., Clouthier, S. C., Doran, J. L., Banser, P. A. & Kay, W. W. *Salmonella* Enteritidis *agfBAC* operon encoding thin, aggregative fimbriae. *J. Bacteriol.* **178**, 662–667 (1996).
50. White, A., Gibson, D., Collinson, S., Banser, P. & Kay, W. Extracellular polysaccharides associated with thin aggregative fimbriae of *Salmonella enterica* serovar Enteritidis. *J. Bacteriol.* **185**, 5398–5407 (2003).
51. Römling, U., Bian, Z., Hammar, M., Sierralta, W. D. & Normark, S. Curli fibers are highly conserved between *Salmonella* Typhimurium and *Escherichia coli* with respect to operon structure and regulation. *J. Bacteriol.* **180**, 722–731 (1998).
52. Doran, J. *et al.* DNA-based diagnostic tests for *Salmonella* species targeting *agfA*, the structural gene for thin, aggregative fimbriae. *J. Clin. Microbiol.* **31**, 2263–2273 (1993).
53. Boyd, E. F. & Hartl, D. L. *Salmonella* virulence plasmid: modular acquisition of the *spv* virulence region by an F-plasmid in *Salmonella enterica* subspecies I and insertion into the chromosome of subspecies II, IIIa, IV and VII isolates. *Genetics* **149**, 1183–1190 (1998).
54. Mazurkiewicz, P. *et al.* *SpvC* is a *Salmonella* effector with phosphothreonine lyase activity on host mitogen-activated protein kinases. *Mol. Microbiol.* **67**, 1371–1383 (2008).
55. Matsui, H. *et al.* Virulence plasmid-borne *spvB* and *spvC* genes can replace the 90-kilobase plasmid in conferring virulence to *Salmonella enterica* serovar Typhimurium in subcutaneously inoculated mice. *J. Bacteriol.* **183**, 4652–4658 (2001).
56. Roudier, C., Fierer, J. & Guiney, D. Characterization of translation termination mutations in the *spv* operon of the *Salmonella* virulence plasmid pSDL2. *J. Bacteriol.* **174**, 6418–6423 (1992).
57. Hu, L. *et al.* DNA sequences and predicted protein structures of *prot6E* and *sefA* genes for *Salmonella* ser. Enteritidis detection. *Food Control* **96**, 271–280 (2019).
58. Oriero, E. C., Okebe, J., Jacobs, J., Nwakanma, D. & D'Alessandro, U. Diagnostic performance of a novel loop-mediated isothermal amplification (LAMP) assay targeting the apicoplast genome for malaria diagnosis in a field setting in sub-Saharan Africa. *Malar. J.* **14**, 1–6 (2015).
59. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2-approximately maximum-likelihood trees for large alignments. *PLoS One* **5**, e9490 (2010).
60. Waterhouse, A. *et al.* SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* (2018).
61. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10**, 845 (2015).

Author contributions

Conceived and designed the study: G.Z., L.H., E.W.B., and M.W.A. Performed experiments: L.H. and G.Z. Analyzed data L.H. and G.C. Contributed to the writing of the manuscript: L.H., G.Z., and L.M.

Disclaimer

Mention of trade names or commercial products in the paper is solely for the purpose of providing scientific information and does not imply recommendation or endorsement by the U. S. Food and Drug Administration.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-00224-7>.

Correspondence and requests for materials should be addressed to G.Z.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021