Received: 3 February 2021 | Revised: 7 June 2021 | Accepted: 29 June 2021

DOI: 10.1002/jcla.23901

RESEARCH ARTICLE

WILEY

Multilocus sequence typing analysis and second-generation sequencing analysis of *Salmonella Wandsworth*

Chunling Zhang[1] | **Yichao Wang**[1] | **Yutian Lu**[1] | **Xuebin Xu**[2]

[1]Department of Clinical Laboratory, Taizhou Central Hospital (Taizhou University Hospital), Taizhou, China

[2]Shanghai Municipal Center for Disease Control and Prevention, Shanghai, China

Correspondence
Chunling Zhang and Yichao Wang, Department of Clinical Laboratory, Taizhou Central Hospital (Taizhou University Hospital), Taizhou, Zhejiang, China.
Emails: zhangcl3444@tzzxyy.com; wangyichao9353@tzzxyy.com

Funding information
The research was supported by the Zhejiang Medical and Health Science and Technology Program (2019KY788)

Abstract

Background: *Salmonella Wandsworth* is a rare serotype of *Salmonella*. This study analyzed the genotyping, genome structure, and molecular biological functions of *Salmonella Wandsworth* based on the results of multilocus sequence typing and next-generation sequencing genome assembly analysis.

Methods: Serological typing was performed using the slide-agglutination method. The micro broth dilution method was used to test antibiotic susceptibility. Multilocus sequence typing (MLST) was used to perform the homology analysis, while the second-generation sequencing genome analysis was used to analyze the whole genome of the bacteria.

Results: *Salmonella Wandsworth* is Group Q *Salmonella*. The MLST of this strain was ST1498. *Salmonella Wandsworth* was sensitive to antibiotics, such as ceftriaxone, imipenem, chloramphenicol, and colistin, but was resistant to ampicillin, cefalotin, gentamicin, and ciprofloxacin. The second-generation sequencing results showed that the genome sequence length of the bacteria was 5109457bp. Annotated COG library analysis generated 3,746 corresponding genes. After the comparison with the KEGG library, 1,340 genes, which participate in 19 types of metabolic pathways, were obtained. A total of 249 pathogenic factors and 2 disease islands were predicted. 2 CRISPR sites and 8 Cas sites were predicted. It can be seen from the evolutionary tree that *Salmonella Wandsworth MLST1498* and *Paratyphi B str.SPB7* are gathered together. We identified one resistance gene, namely, aac(6')-Iaa accounting for aminoglycoside resistance.

Conclusion: *Salmonella Wandsworth* isolated in this study is *Salmonella* group Q. Consequently, it is necessary to strengthen the understanding of clinical infections of *Salmonella Wandsworth* and carry out continuous monitoring and research.

KEYWORDS
antibiotic resistance, Multilocus sequence typing (MLST), next-generation sequencing, *Salmonella Wandsworth*, serotype

1 | INTRODUCTION

Salmonella is an important foodborne pathogen that causes acute diarrhea in humans worldwide.[1] In approximately 5%–10% of patients, the condition is further complicated by bacteremia, which can be life threatening.

According to the serotype classification scheme proposed by Kauffman-White,[2] *Salmonella* can be divided into A, B, C1, C2, D, E ~Z, and other serogroups. These serogroups can be further subdivided into 2,579 serotypes. The main basis for serotype classification are "O" antigen, "H" antigen and "Vi" antigen. The "O" antigens are the polysaccharides associated with the lipopolysaccharide of the bacterial outer membrane. The "H" antigens are proteins associated with the bacterial flagella. Salmonellas exist in two phases: a motile phase and a nonmotile phase. These are also referred to as the specific and nonspecific phases. "Vi" antigen is named because it encodes a virulence-related gene, which is only present in the three serotypes of *Salmonella typhimurium*, *Salmonella paratyphoid C*, and *Salmonella Dublin*. The "O" antigen of *Salmonella Wandsworth* is O:39. The "H" antigens are H1:b and H2:1,2. According to the serotype classification scheme proposed by Kauffman-White, *Salmonella Wandsworth* belongs to *Salmonella* serogroup Q. Most strains come from farmed freshwater and marine products, especially amphibians, such as bullfrogs and soft-shelled turtles. The ecological cycle of *Salmonella Wandsworth* does exist between farmers and snakes in Thai snake farms.[3]

In 2009, an outbreak of infection by *Salmonella Wandsworth* serotype was reported in the United States.[4] Over recent years, this bacterium has been rarely reported. Studies suggest that the infection of *Salmonella Wandsworth* is mainly related to wild-caught and farm-raised animals.[3,5]

Given that the existing research on *Salmonella Wandsworth* has been mainly based on the source analysis, genotyping and next-generation sequencing research are rare.[6] Therefore, we performed multilocus sequence typing analysis and next-generation sequencing genome assembly analysis on a strain of *Salmonella Wandsworth* isolated from a patient with clinical blood-borne infections. The results of the study are reported as follows.

2 | MATERIALS AND METHODS

2.1 | Collection of clinical specimens

The strain was isolated from a blood culture specimen of a 10-year-old child with a history of multiple convulsions, who was admitted to the hospital with gastroenteritis and systemic inflammatory response syndrome. The blood samples and fecal samples of the patients were collected and cultured. The patient provided the written informed consent.

This study was approved by the Ethics committee of Taizhou Central Hospital, Affiliated Hospital of Taizhou University.

2.2 | Identification of bacterial strains

VITEK 2 compact and matrix-assisted laser analysis time-of-flight mass spectrometry was used for identification. The serotype identification of the *Salmonella* serotype provided by Ningbo Tianrun Company was used to determine the strain's serotype.

2.3 | Antimicrobial sensitivity test

The micro broth dilution method and the disk diffusion method recommended by the American Committee for Clinical and Laboratory Standards (CLSI) was used for antimicrobial susceptibility testing. Antibacterial drugs included ampicillin, cefalotin, cefoxitin, ceftriaxone, ceftazidime, imipenem, ertapenem, ciprofloxacin, tetracycline, colistin, chloramphenicol, piperacillin/tazobactam, gentamicin, apramycin, colistin, and trimethoprim/sulfamethoxazole (SXT1.25/23.75ug). The results were interpreted in accordance with M100 performance standards for antimicrobial susceptibility testing (CLSI-M100), and *Escherichia coli* ATCC25922 was used as the quality control bacteria for the antimicrobial susceptibility test.

2.4 | Multilocus sequence typing

The homology analysis of this strain was carried out by using multilocus sequence typing (MLST). First, the *Salmonella* genome was extracted, and seven *Salmonella* housekeeping genes were selected (*aroC*, *dnaN*, *thrA*, *hemD*, *sucA*, *hisD*, *purE*) according to the database PubMLST for PCR amplification. The Hangzhou Golden Field Medical Inspection Co.,Ltd. was entrusted to carry out the amplification products sequencing. The sequencing result was compared with the standard sequence and was uploaded to the database, and the final sequencing result was compared with the MLST official website to obtain the corresponding allelic profile and sequence typing number (ST) of seven pairs of housekeeping genes.

2.5 | Next-generation sequencing analysis

Zhejiang Tianke High-tech Development Co., Ltd. was entrusted to perform the second-generation sequencing genome analysis of this strain of *Salmonella Wandsworth*. Assembly analysis and gene prediction were performed. COG and KEGG databases were used to predict and analyze gene function, VFDB database was used to predict bacterial virulence factors, and CRISPRC was used as finder software to predict the CRISP/Cas system in the bacterial genome. *Salmonella* phylogenetic tree was constructed based on the whole- genome sequence. The AAF (alignment and assembly-free, https://github.com/fanhuan/AAF) method was used to construct a genomic evolutionary tree. ResFinder and PlasmidFinder databases were used to analyze the antibiotic resistance genes and the plasmid types, respectively.

3 | RESULTS

3.1 | Strain identification results

The bacterial identification revealed *Salmonella Wandsworth*, which was classified as *Salmonella* group Q.

3.2 | Antimicrobial susceptibility test results

The results of the antimicrobial susceptibility test showed that *Salmonella Wandsworth* was sensitive to antibiotics, such as ceftriaxone, ceftazidime, ertapenem, imipenem, piperacillin/tazobactam, chloramphenicol, and colistin,but was resistant to ampicillin, cefalotin, cefoxitin, trimethoprim/sulfamethoxazole, gentamicin, apramycin,tetracycline, and ciprofloxacin.

3.3 | Results of Multilocus sequence typing

The PCR amplification products of seven housekeeping genes of *Salmonella Wandsworth (aroC*, *dnaN*, *hemD*, *hisD*, *purE*, *sucA*, *and thrA)* were detected by agarose gel electrophoresis. The electrophoresis pattern results were consistent with the expected results, and the bands were clear. The allele numbers of the seven housekeeping genes of this strain of *Salmonella* were 14, 13, 43, 17, 96, 19, and 17, and the corresponding sequence type was ST1498.

3.4 | Basic information of genome sequence

The Unicycler software was used for assembly, which revealed that the genome of this strain of *Salmonella Wandsworth* was 5109457bp with a GC content of 52%. Glimmer 3.02 software predicted 5,412 coding genes, with a total length of 4420352bp, and an average length of 816.77bp, accounting for 86.51% of the total length of the genome. Seventy-eight tRNAs and 3 rRNAs were predicted.

3.5 | COG prediction analysis of gene group

Comparison with the COG database resulted in 4,727 related genes. The gene information was classified according to the COG classification standard. The genome was divided into 23 categories: RNA processing and modification (A), a total of 2 genes; Chromatin structure and dynamics (B), a total of 1 gene; Energy production and conversion (C), a total of 314 genes; Cell cycle control, cell division, chromosome partitioning (D), a total of 40 genes; Amino acid transport and metabolism (E), 466 genes in total; Nucleotide metabolic transport (F), 95 genes in total; Carbohydrate transport and metabolism (G), 432 genes in total; Coenzyme transport and metabolism (H), 181 genes in total; Lipid transport and metabolism (I), a total of 104 genes; Translation, ribosomal structure and biogenesis (J), a total of 184 genes; Transcription (K), a total of 377 genes; Replication, recombination and repair (L), a total of 228 genes; Cell wall/membrane/envelope biogenesis (M), a total of 254 genes; Cell motility (N), a total of 130 genes; Posttranslational modification, protein turnover, chaperones (O), a total of 159 genes; Inorganic ion transport and metabolism (P), a total of 290 genes; Secondary metabolites biosynthesis, transport and catabolism (Q), a total of 84 genes; General function prediction only (R), 575 genes; Function unknown (S), 377 genes; Signal transduction mechanisms (T), a total of 237 genes; Intracellular trafficking, secretion, and vesicular transport (U), a total of 144 genes; Defense mechanisms (V), a total of 52 genes; Extracellular structures (W), a total of 1 gene (Figure 1).

3.6 | Classification of KEGG metabolic pathways of the genome

In this study, we compared the KEGG library with 1,340 corresponding genes and divided them into 19 types of metabolic pathways, which correspond to the expression products of different gene sequences. The function and quantity of each type of gene are shown in Figure 2. Replication and repair, a total of 60 genes; Transcription, a total of 16 genes; Translation, a total of 71 genes; Folding, sorting, and degradation, a total of 49 genes; Transport and catabolism, a total of 28 genes; Signal transduction, a total of 4 genes; Membrane transport, a total of 36 genes; Infectious diseases: Parasitic, a total of 17 genes; Metabolism of cofactors and vitamins, a total of 138 genes; Amino acid metabolism, a total of 402 genes; Nucleotide metabolism, a total of 136 genes; Metabolism of terpenoids and polyketides, a total of 15 genes; Global map, a total of 975 genes; Glycan biosynthesis and metabolism, a total of 15 genes; Lipid metabolism, a total of 157 genes; Energy metabolism, a total of 92 genes; Carbohydrate metabolism, a total of 497 genes; Metabolism of other amino acids, a total of 86 genes; Biosynthesis of other secondary metabolites, a total of 4 genes.

3.7 | Pathogenic virulence factor of *Salmonella Wandsworth*

BLAST software was used to compare the amino acid sequence of the target species with the VFDB database (E-value<0.00001) and combine the gene of the target species with its corresponding functional annotation information to obtain the annotation result. Since there may be more than one sequence alignment result for each sequence, to ensure biological significance, the screening conditions for annotation required to retain an alignment result with the highest score and credibility greater than 80% as the annotation of the gene.

A total of 249 pathogenic factors and 2 pathogenic islands were predicted. They mainly included 9 categories, such as iron uptake,

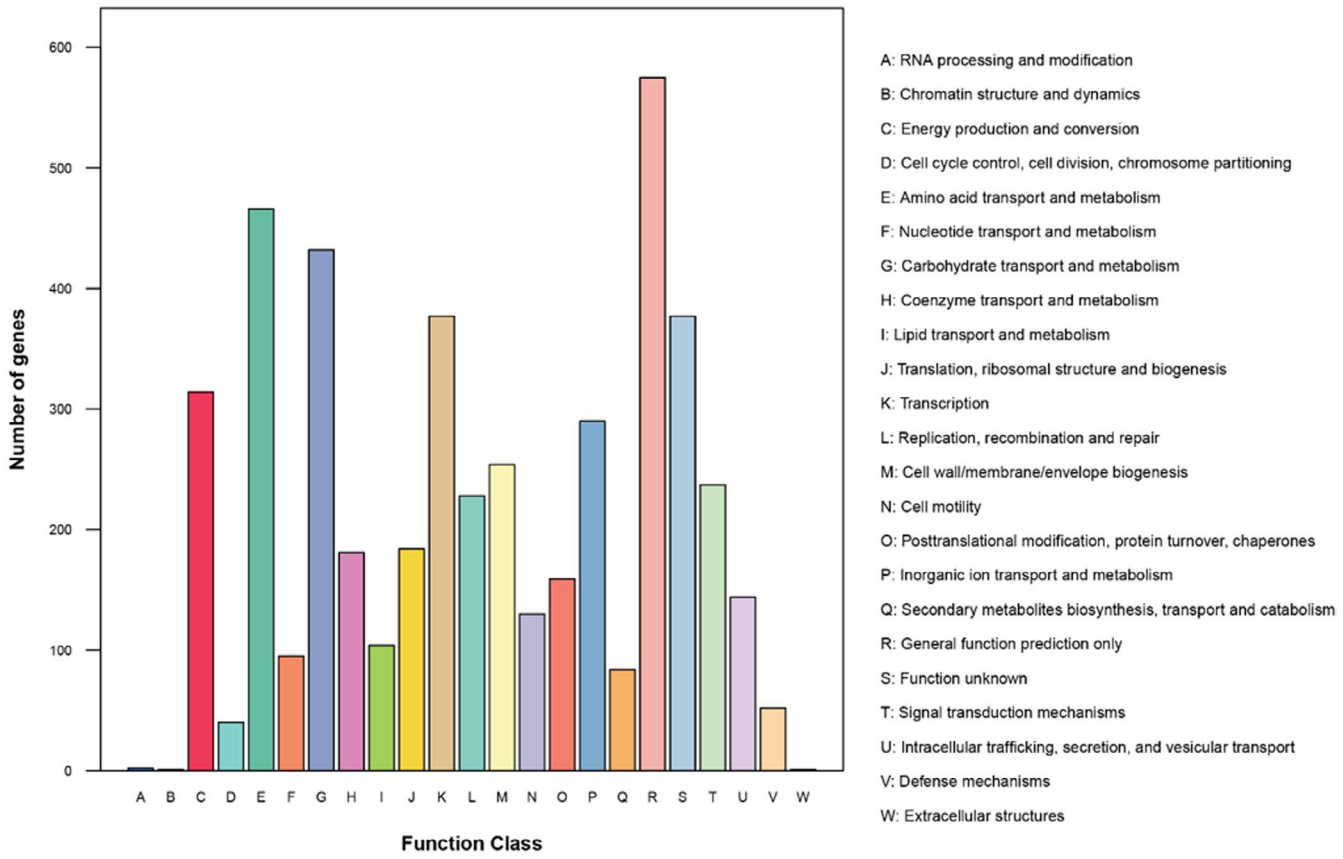

FIGURE 1 COG annotated classification chart of Salmonella Wandsworth

invasion, adhesion, antiphagocytosis, endotoxin, secretion system, toxin, and regulatory gene (Figure 3).

3.8 | CRISPR_Cas prediction of *Salmonella Wandsworth*

The CrisprCas finder software was used to predict 3 CRISP antiphagocytosis, R repeat sequences and 8 cascluster sequences in the genome of this strain. Among them, contig_2_len_632071_1 started at 451157bp and ended at 452710bp, reaching a total of 1553bp. Repeat sequence GTGTTCCCCGCGCCAGCGGGGATAAACCG. contig_2_len_632071_2 started at 469006bp and ended at 470193bp, totaling 1187bp. Repeat sequence GTGTTCCCCGCG CCAGCGGGGATAAACCG. contig_16_len_92086_1 started at 5534bp and ended at 5623bp, totaling 89bp. Repeat sequence GTATCTGGTAACCGGCTGACCAGTCTGCCG. See Table 1 for details.

3.9 | *Salmonella* phylogenetic tree constructed based on the whole-genome sequence

At present, there are 15 strains *S. enterica* subsp.I with complete genome sequences. Based on these genome sequences, the AAF method was used to construct a genomic evolutionary tree. ((Figure 4)) It can be seen from the evolutionary tree that *Salmonella Wandsworth MLST1498* and *Paratyphi B str.SPB7* are gathered together, and the evolutionary distance between the two is very short.

3.10 | Antibiotic resistance genes and plasmids content analysis

We identified one resistance gene, namely, aac(6')-Iaa accounting for aminoglycoside resistance using ResFinder databases. Position in contig is 135553–135990. The assembly sequence was analyzed by the PlasmidFinder software and a suspected IncFII(S) plasmid fragment was found.

4 | DISCUSSION

The *Salmonella Wandsworth* detected in this study is a relatively rare serotype. Its O antigen is outside the AF group and belongs to *Salmonella* group Q.[7,8] Currently, there are only sporadic reports on the clinically isolated *Salmonella Wandsworth*. The serotyping of the strain in this study was entrusted to the Shanghai Center for Disease Control and Prevention, and the serotyping of *Salmonella Wandsworth* was

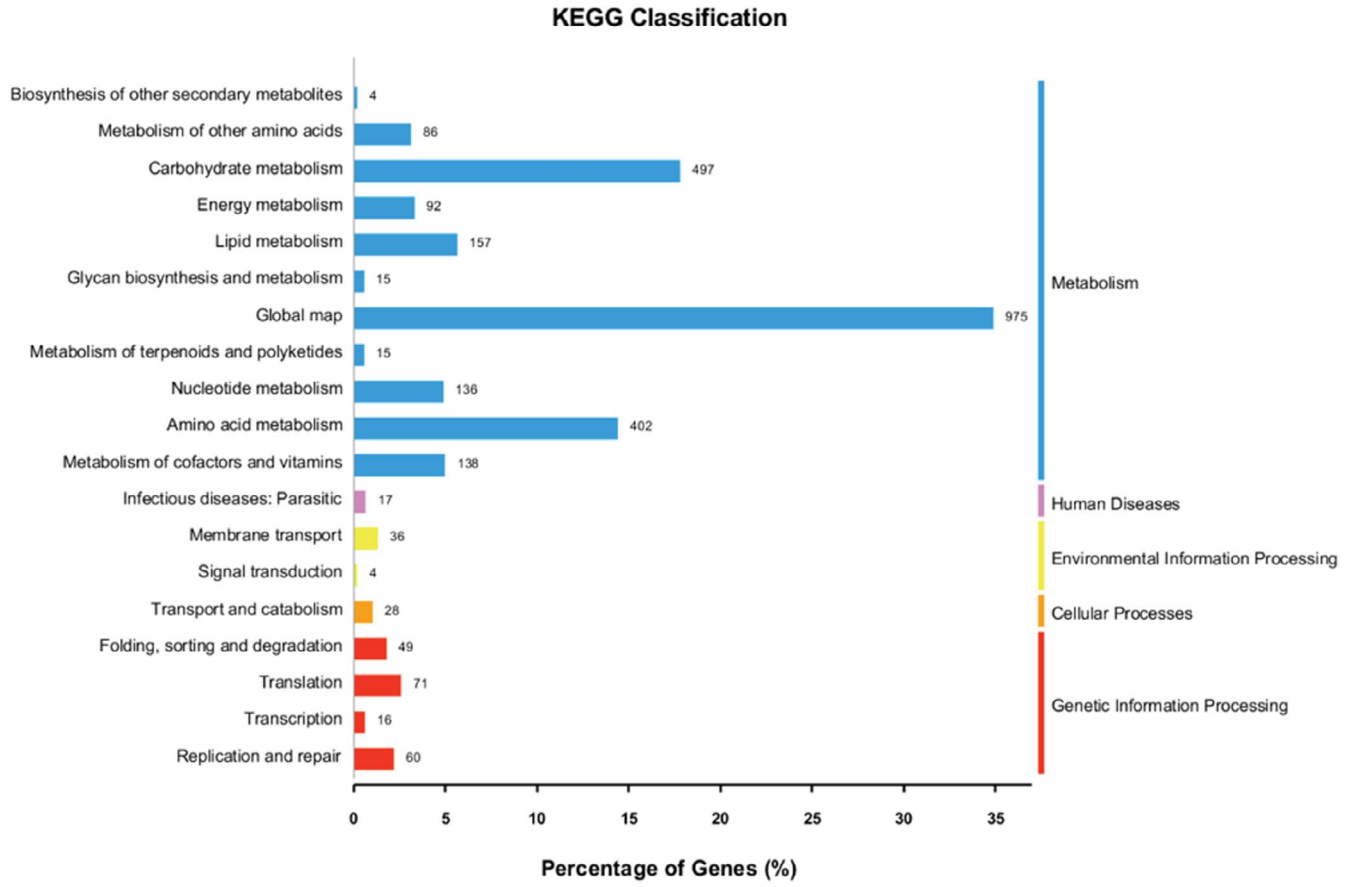

FIGURE 2 KEGG pathway classification for genes of Salmonella Wandsworth. Note: The ordinate indicates the name of the KEGG metabolic pathway, and the abscissa indicates the number of genes annotated to the pathway and its proportion to the total number of genes annotated

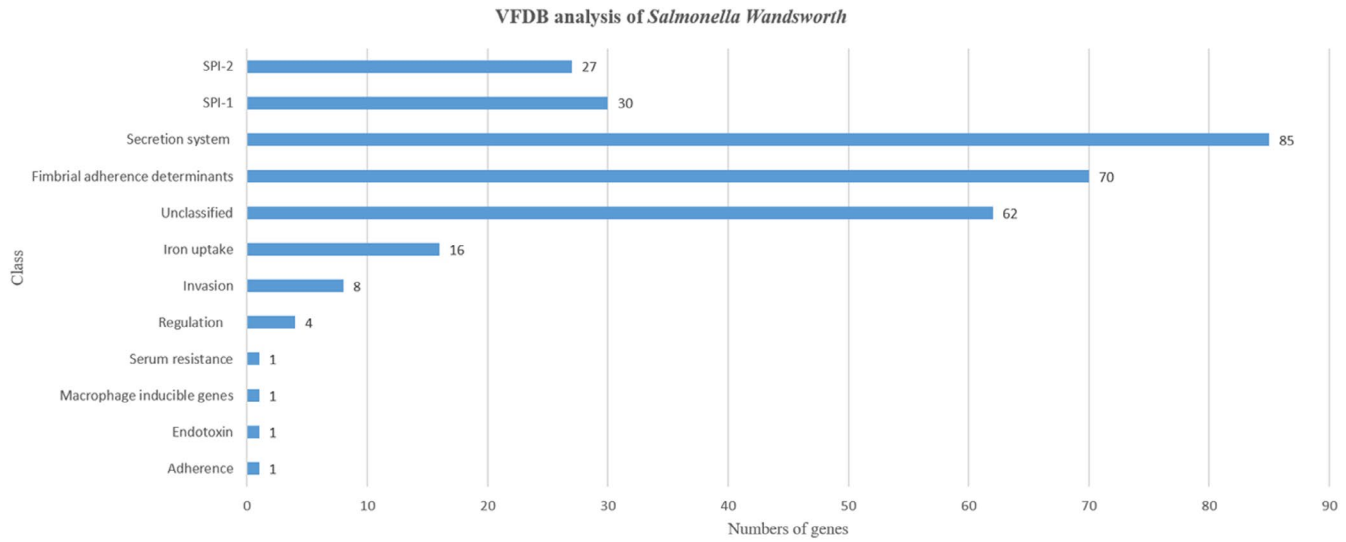

FIGURE 3 VFDB analysis of Salmonella Wandsworth. Note: The ordinate indicates classification of virulence factor, and the abscissa indicates the numbers of Virulence gene

completed. There are not many reports about *Salmonella Wandsworth*, which may also be due to the lack of diagnostic serum in primary hospitals. Therefore, primary laboratories should also pay more attention to the isolation and identification of non-A-F *Salmonella*.

Non-typhoid *Salmonella* mainly causes self-limiting enteritis, while a small part can also cause sepsis and be life-threatening.[9] In this study, the patient with bloodstream infection of *Salmonella Wandsworth* was a 10-year-old child with a history of

TABLE 1 CRISPR-Cas Statistical interpretation of *Salmonella Wandsworth*

Sequence	Element	Start	End	Spacer/gene	Repeat consensus
Contig1	CRISPR	451157	452710	GTGTTCCCCGC GCCAGCGGGG ATAAACCG	2
contig2_	Cas cluster			Cas3 Cas1 Cas2 Cas7 Cas5 Cas6	8
Contig3	CRISPR	5534	5623	GTATCTGGTAACC GGCTGACCAGTCT GCCG	1

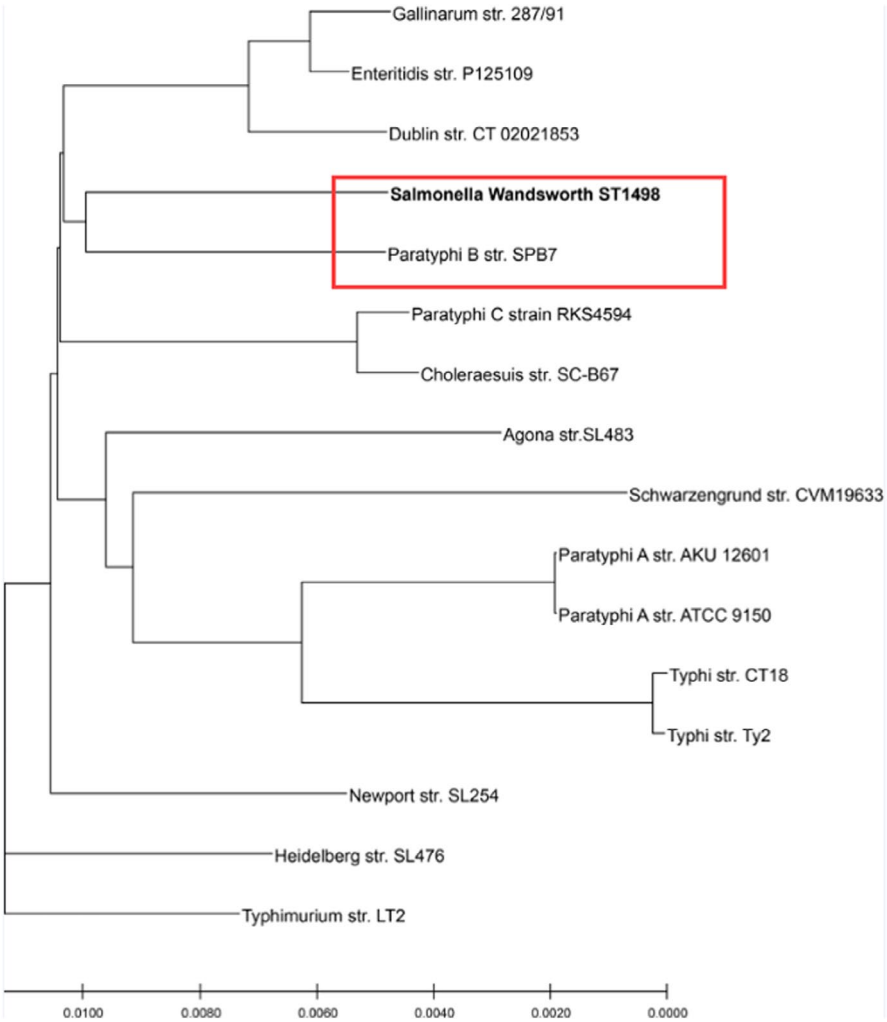

FIGURE 4 Salmonella phylogenetic tree constructed based on the whole-genome sequence

hypoimmunoglobulinemia, pneumonia, encephalitis, etc. *Salmonella Wandsworth* was isolated from blood culture. Considering the common colonization and infection sites of the bacteria are intestines, it remains unclear how the bacteria get into the blood. Accordingly, the microbiology laboratory contacted the clinic suggesting fecal culture; however, *Salmonella Wandsworth* could not be found from the fecal culture. At the same time, as there was no further investigation and research on the living environment of the patient, there was no more data to support the traceability of the bloodstream infection of *Salmonella Wandsworth*.

In this study, *Salmonella Wandsworth* was sensitive to ceftriaxone, ceftazidime, ertapenem, imipenem, piperacillin/tazobactam, chloramphenicol, colistin, while it was resistant to ampicillin, cefalotin, cefoxitin, trimethoprim/sulfamethoxazole, gentamicin,

apramycin,tetracycline, and ciprofloxacin. Compared with previous reports [10,11] the strain in this study is highly resistant to gentamicin. At the same time, the strain's resistance gene analysis showed that one resistance gene namely aac(6')-Iaa accounting for aminoglycoside resistance was found using ResFinder databases.

At present, the development of whole-genome sequencing technology is advancing very rapidly[12] laying the foundation for the research from the level of individual genes to the genome's overall function.[13] However, there is no research on the genome of *Salmonella Wandsworth*. In this study, the complete genome of *Salmonella Wandsworth* was obtained by performing whole-genome sequencing and bioinformatics analysis on clinical isolates of *Salmonella Wandsworth*. The obtained results revealed the length of 5109457bp and the GC content of 52%.

The biological functions of the bacteria were analyzed using biological software, and 4,727 genes were obtained through COG library annotation. These genes are involved in the synthesis and metabolism processes to varying degrees. The transportation of compounds and amino acids are closely related to the toxicity of *Salmonella Wandsworth*.[14] By comparing the KEGG library, the metabolic pathways were significantly enriched in amino acids, membrane operation, energy, carbohydrates, etc. These important functions or enrichment pathways are related to the regulation of intracellular transport, virulence, and gene expression.[15,16]

CRISP is a series of short palindromes with many clusters and regular intervals.[17] CRISP has an immune role in prokaryotes, can resist foreign plasmid and phage sequences, and can identify and silence invading functional elements.[18] In this study, CrisprCas finder software was used to predict 3 CRISPR repeat sequences and eight cascluster sequences in the genome of this strain. At the same time, this study also used multilocus sequence typing technology (MLST) to determine the seven housekeeping genes of *Salmonella Wandsworth*. The corresponding sequence type was ST1498,which was rare genotype. In the study by scholars of Tehran Hospitals,[10,11] the sequence type of *Salmonella* is mainly ST32 and ST19. Based on CRISP and MLST genotyping techniques, *Salmonella Wandsworth* was genotyped, laying a foundation for the evolution of *Salmonella Wandsworth*.

In this study, 249 pathogenic factors, 2 pathogenic islands, and related virulence genes related to invasion, adhesion, antiphagocytosis, endotoxin, and type III secretion system were obtained through the analysis of the VFDB database. The number of TTSS virulence genes in the Type III secretion system resulted as the database which can directly or indirectly affect the metabolism and transport system of the organism.

It can be seen from the evolutionary tree that *Salmonella Wandsworth MLST1498* and *Paratyphi B str.SPB7* are gathered together, and the evolutionary distance between the two is very short, similar to the evolutionary distance between different strains in the same serotype, such as *Typhi str.CT18* and *Typhi. str. Ty2*. The host of *Salmonella paratyphi B str. SPB7* infection is most common in children and adolescents. The close genetic relationship between *Salmonella Wandsworth ST1498* and *Salmonella paratyphi B str. SPB7* indicates that the host-related traits of *Salmonella Wandsworth* may enter the genome in a convergent evolutionary manner, gradually transforming from cold-blooded animal origin to human origin.

The assembly sequence was analyzed by the PlasmidFinder software and a suspected IncFII(S) plasmid fragment was found. Because it is a draft data, the genome is relatively fragmented, and the fragment size is only 262 bases, which is relatively small. But the IncFII(S)_CP000858 plasmid is about 55Kb in size. Later, we will continue to complete the map of the bacteria to clarify the plasmid.

Through second-generation sequencing and multi-site sequence typing, this study deeply analyzed the genotyping and biological function prediction of the clinical isolates of *Salmonella Wandsworth*, which is a valuable method for the development and diagnosis of *Salmonella Wandsworth* vaccine. The establishment and subsequent experimental research are warranted to provide molecular basis and guidance.

DATA AVAILABILITY STATEMENT

All data relevant to the study are included in the article or uploaded as supplementary information.

ORCID

Chunling Zhang https://orcid.org/0000-0001-8607-9669

REFERENCES

1. Yue M, Li X, Liu D, Hu X. Serotypes, antibiotic resistance, and virulence genes of Salmonella in children with diarrhea. *J Clin Lab Anal*. 2020;34(12):e23525.
2. Le Minor L, Bockemuhl J. [1987 supplement (no. 31) to the schema of Kauffmann-White]. *Ann Inst Pasteur Microbiol*. 1988;139(3):331-335.
3. Prapasarakul N, Pulsrikarn C, Vasaruchapong T, et al. Salmonella serovar distribution in cobras (Naja kaouthia), snake-food species, and farm workers at Queen Saovabha Snake Park, Thailand. *J Vet Diagn Invest*. 2012;24(2):288-294.
4. Sotir MJ, Ewald G, Kimura AC, et al. Outbreak of Salmonella Wandsworth and typhimurium infections in infants and toddlers traced to a commercial vegetable-coated snack food. *Pediatr Infect Dis J*. 2009;28(12):1041-1046.
5. Ribas A, Poonlaphdecha S. Wild-caught and farm-reared amphibians are important reservoirs of Salmonella, a study in North-East Thailand. *Zoonoses Public Health*. 2017;64(2):106-110.
6. Zhang J, Yang X, Kuang D, et al. Prevalence of antimicrobial resistance of non-typhoidal Salmonella serovars in retail aquaculture products. *Int J Food Microbiol*. 2015;210:47-52.
7. Ling J, Chau PY, Rowe B. Salmonella serotypes and incidence of multiply-resistant Salmonellae isolated from diarrhoeal patients in Hong Kong from 1973–82. *Epidemiol Infect*. 1987;99(2):295-306.
8. Reilly PJ, Twiddy DR. Salmonella and vibrio cholerae in brackishwater cultured tropical prawns. *Int J Food Microbiol*. 1992;16(4):293-301.
9. Farmakiotis D, Ntouloulis C, Mihailidis N, et al. Diagnostic value of eosinopenia in non-Typhi Salmonella enteritis. *Clin Infect Dis*. 2014;59(8):1197-1198.
10. Ranjbar R, Elhaghi P, Shokoohizadeh L. Multilocus sequence typing of the clinical isolates of Salmonella enterica serovar typhimurium in Tehran hospitals. *Iran J Med Sci*. 2017;42(5):443-448.
11. Ranjbar R, Rahmati H, Shokoohizadeh L. Detection of common clones of Salmonella enterica serotype Infantis from human

sources in Tehran hospitals. *Gastroenterol Hepatol Bed Bench*. 2018;11(1):54-59.
12. Lappalainen T, Scott AJ, Brandt M, Hall IM. Genomic analysis in the age of human genome sequencing. *Cell*. 2019;177(1):70-84.
13. Sohn JI, Nam JW. The present and future of de novo whole-genome assembly. *Brief Bioinform*. 2018;19(1):23-40.
14. Galperin MY, Makarova KS, Wolf YI, Koonin EV. Expanded microbial genome coverage and improved protein family annotation in the COG database. *Nucleic Acids Res*. 2015;43(Database issue):D261-D269.
15. Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res*. 2017;45(D1):D353-D361.
16. Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res*. 2016;44(D1):D457-D462.
17. Weigel Munoz M, Carvajal G, Curci L, Gonzalez SN, Cuasnicu PS. Relevance of CRISP proteins for epididymal physiology, fertilization, and fertility. *Andrology*. 2019;7(5):610-617.
18. Cheng S. The CRISP theory of hippocampal function in episodic memory. *Front Neural Circuits*. 2013;7:88.

How to cite this article: Zhang C, Wang Y, Lu Y, Xu X. Multilocus sequence typing analysis and second-generation sequencing analysis of *Salmonella Wandsworth*. *J Clin Lab Anal*. 2021;35:e23901. https://doi.org/10.1002/jcla.23901