

# *Salmonella* serotypes in the genomic era: simplified *Salmonella* serotype interpretation from DNA sequence data

Xiangyu Deng,<sup>1</sup> Shaoting Li,<sup>1,2</sup> Tongzhou Xu,<sup>1</sup> Zhemin Zhou,<sup>3</sup> Michelle M. Moore,<sup>4</sup> Ruth Timme,<sup>5</sup> Shaohua Zhao,<sup>6</sup> Charlotte Lane,<sup>7</sup> Blake A. Dinsmore,<sup>7</sup> François-Xavier Weill,<sup>8</sup> Patricia I. Fields<sup>7</sup>

**AUTHOR AFFILIATIONS** See affiliation list on p. 10.

**ABSTRACT** In the era of genomic characterization of strains for public health microbiology, whole genome sequencing (WGS)-enabled subtyping of *Salmonella* provides superior discrimination of strains compared to traditional methods such as serotyping. Nonetheless, serotypes are still very useful; they maintain historical continuity and facilitate clear communication. Genetic determination of serotypes from WGS data is now routine. Genetic determination of rarer serotypes can be problematic due to a lack of sequences for rare antigen types and alleles, a lack of understanding of the genetic basis for some antigens, or some inconsistencies in the White-Kauffmann-Le Minor (WKL) Scheme for *Salmonella* serotype designation. Here, we present a simplified interpretation of serotypes to address the shortcomings of genetic methods, which will allow the streamlined integration of serotype determination into the WGS workflow. The simplification represents a consensus perspective among major U.S. public health agencies and serves as a WGS-oriented interpretation of the WKL Scheme. We also present SeqSero2S, a bioinformatics tool for WGS-based serotype prediction using the simplified interpretation.

**IMPORTANCE** The utility of *Salmonella* serotyping has evolved from a primary subtyping method, where the need for strain discrimination justified its complexity, to a supplemental subtyping scheme and nomenclature convention, where clarity and simplicity in communication have become important for its continued use. Compared to phenotypic methods like serotyping, whole genome sequencing (WGS)-based subtyping methods excel in recognizing natural populations, which avoids grouping together strains from different genetic backgrounds or splitting genetically related strains into different groups. This simplified interpretation of serotypes addresses a shortcoming of the original scheme by combining some serotypes that are known to be genetically related. Our simplified interpretation of the White-Kauffmann-Le Minor (WKL) Scheme facilitates a complete and smooth transition of serotyping's role, especially from the public health perspective that has been shaped by the routine use of WGS.

**KEYWORDS** *Salmonella*, serotype, whole genome sequencing

Despite substantial progress in foodborne pathogen surveillance and control, *Salmonella* remains a leading cause of foodborne infections in the United States, showing a stalled decline in recent years (1–3). U.S. surveillance of *Salmonella* has been built upon serotypes, also called serovars, a typing scheme established in 1934 (4). Over 90 years of literature and 50 years of U.S. surveillance data based on serotype (5) have contributed to a well-established, serotype-centric understanding of *Salmonella* epidemiology, biology, and evolution. For example, specific serotypes are associated with specific disease syndromes (6), host associations (7), and geographic and/or source

**Editor** Gemma Reguera, Michigan State University, East Lansing, Michigan, USA

Address correspondence to Xiangyu Deng, [xdeng@uga.edu](mailto:xdeng@uga.edu), or Patricia I. Fields, [pif1@cdc.gov](mailto:pif1@cdc.gov).

The authors declare no conflict of interest.

**Received** 31 December 2024

**Accepted** 16 January 2025

**Published** 24 February 2025

Copyright © 2025 Deng et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

distribution (8, 9). Serotypes are integral to *Salmonella* nomenclature and serve as a universal terminology for communicating *Salmonella* surveillance and research.

*Salmonella* serotypes are designated according to the conventions of the White-Kauffmann-Le Minor (WKL) Scheme (10). The scheme is based on the characterization of *Salmonella* strains based on two surface structures, the lipopolysaccharide O antigen and flagellar H antigen. The scheme recognizes 46 major O antigens or serogroups plus several secondary O antigens that are found in specific serogroups. Serogroup antigens are commonly encoded by the *rfb* region on the chromosome and the secondary O antigens that have been characterized are typically encoded by mobile genetic elements. H antigens are proteins that are part of the bacterial flagella. They have been characterized as either single epitopes or multiple epitopes. Multi-epitope H antigens are called H antigen complexes. One example, 1 Complex antigens, share a common epitope "1" and differ by one or more secondary epitopes (e.g., H:1,2; H:1,5; H:1,6; H:1,7; and H:1,5,7). Since the same antigenic types may appear in different taxonomic lineages, subspecies identification is commonly included in serotype designations. More than 2,600 specific combinations of O and H antigens (and sometimes additional markers), each a different serotype (or serovar), were described in the current ninth edition of the scheme, which was published in 2007 and has been updated two times since then (11, 12).

Methods to determine serotypes have evolved from phenotypic determination using antigen-antibody agglutination assays to genetic determination using the genes responsible for serotype (13, 14). Genetic determinants include the *rfb* gene cluster encoding genes involved in the biosynthesis of O group antigen and the *fljC* and *fljB* genes encoding the flagellar antigens. Serotype determination using whole genome sequencing (WGS) data was introduced in 2015 (15, 16). SeqSero2, a tool for predicting serotypes based on the genes responsible for serotype (17), has been validated and is routinely used by the Centers for Disease Control and Prevention (CDC), the Food and Drug Administration (FDA), and the U.S. Department of Agriculture Food Safety and Inspection Service (USDA FSIS). It has been incorporated into NCBI Pathogen Detection (<https://www.ncbi.nlm.nih.gov/pathogens>) and Enterobase (<https://enterobase.warwick.ac.uk/>), and validated for the FDA Bacteriological Analytical Manual (<https://www.fda.gov/food/laboratory-methods-food/foods-program-compendium-analytical-laboratory-methods>).

While serotypes continue to be important for disease surveillance and public health investigations, the antigenic detail in the WKL Scheme is less critical than it once was because serotyping is no longer our primary subtyping method. With the launch of PulseNet in 1996, pulsed-field gel electrophoresis (PFGE) replaced serotyping as the primary subtyping method for major foodborne pathogens (18, 19). Since 2013 (20), PFGE was gradually superseded by WGS-based subtyping methods, such as methods based on single nucleotide polymorphisms (SNP) (21) and genome-wide multi-locus sequence typing in outbreak detection and investigation, e.g., core genome multilocus sequence typing (cgMLST) (22).

Several challenges complicate the replication of a scheme based on phenotype into a genetic one. Although the genetic basis for many serotypes is well understood, it is unknown for some antigenic types described in the WKL Scheme. Genetic methods such as SeqSero2 and SISTR, a combination of *in silico* serotyping and 7-gene multilocus sequence typing (MLST) (14), may incorrectly call an antigenic type when a close allele is not present in the database. Furthermore, some serotypes in the scheme are based on phenotypic characteristics beyond O and H antigens, and inconsistencies in how variable epitopes are presented in the scheme make it difficult to assign some serotypes.

While it is currently impossible to replicate the phenotypic serotype scheme with genetic methods, it is also unnecessary because highly discriminatory genetic methods are now available to provide differentiation beyond serotype data. Genetic methods better reveal natural population groups, which is fundamental for surveillance. Dropping or combining certain serotypes, such as those based on unstable antigens, for instance, O and H antigens that may be present or absent without relation to phage conversion

(designated by square brackets in the WKL Scheme), makes serotypes better represent natural populations. Also, the primary value of serotype data in the modern world—clear communication and historical continuity—is diminished for many of the 2,600+ serotypes that are quite rare. Their names are not familiar to the scientific community and defining biologic or epidemiologic characteristics have not been described. A consensus has been reached among a consortium of CDC, FDA, USDA, and SeqSero2 developers to design and implement a simplified interpretation of the scheme for better and more consistent sharing and communication of *Salmonella* serotypes. The goal of the simplification is a scheme where serotype designations can be determined and validated using currently available sequence data and genetic methods. The simplification is based on our experience of determining, communicating, and curating serotypes in public health settings in the United States. Some of the proposed simplifications are already practiced by U.S. agencies. The simplified interpretation has little to no impact on commonly encountered serotypes in the United States and will allow the full transition of serotype determination from a phenotypic to a genetic method. We provide an implementation of the simplified interpretation, designated as SeqSero2S (<https://www.denglab.info/SeqSero2>, <https://github.com/denglab/SeqSero2S>), including additional functionalities to improve genomic prediction of serotypes in general and mitigate potential drawbacks of the simplified scheme.

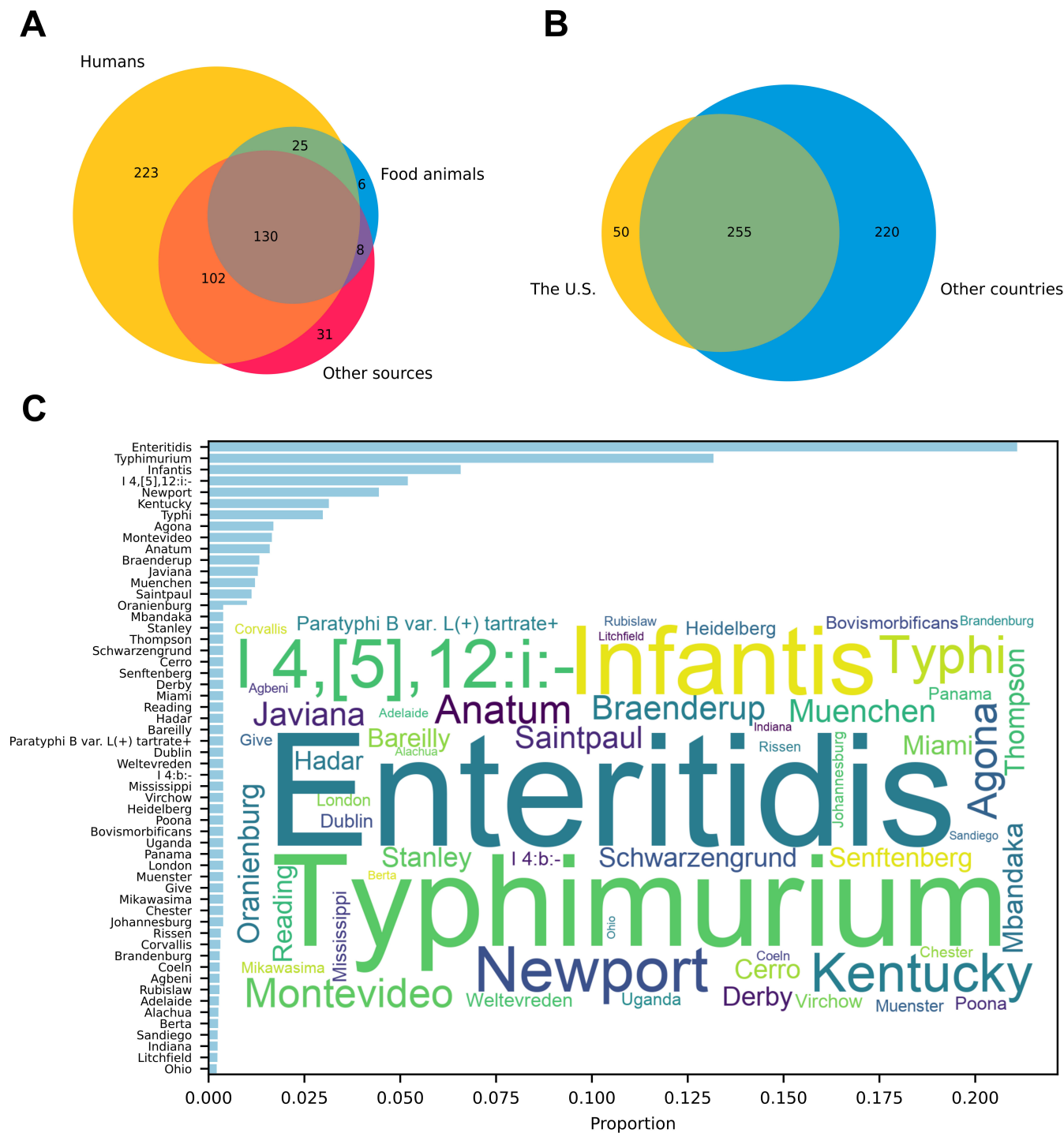
## RESULTS

### Currently circulating *Salmonella* serotypes

We surveyed genomes deposited in NCBI Pathogen Detection and Enterobase between 2018 and 2023. The genomes ( $n = 90,863$ ) represented 525 serotypes from humans, major food animals (poultry, bovine, and swine), and other sources (e.g., environment) were reported worldwide from 2018 to 2023 (Fig. 1A; Table S3). Among these serotypes, each of which included at least two genomes in our data sets, 305 were reported in the United States, 475 were reported from 108 other countries, and 255 were shared between the United States and other countries (Fig. 1B). More serotypes were identified in human isolates ( $n = 480$ ) than in isolates recovered from food animals ( $n = 169$ ) and other sources ( $n = 271$ ), including 223 that were only seen among human isolates (Fig. 1A). Worldwide, the 56 most reported serotypes (Fig. 1C) account for 90% of surveyed isolates (Fig. 1C; Table S3).

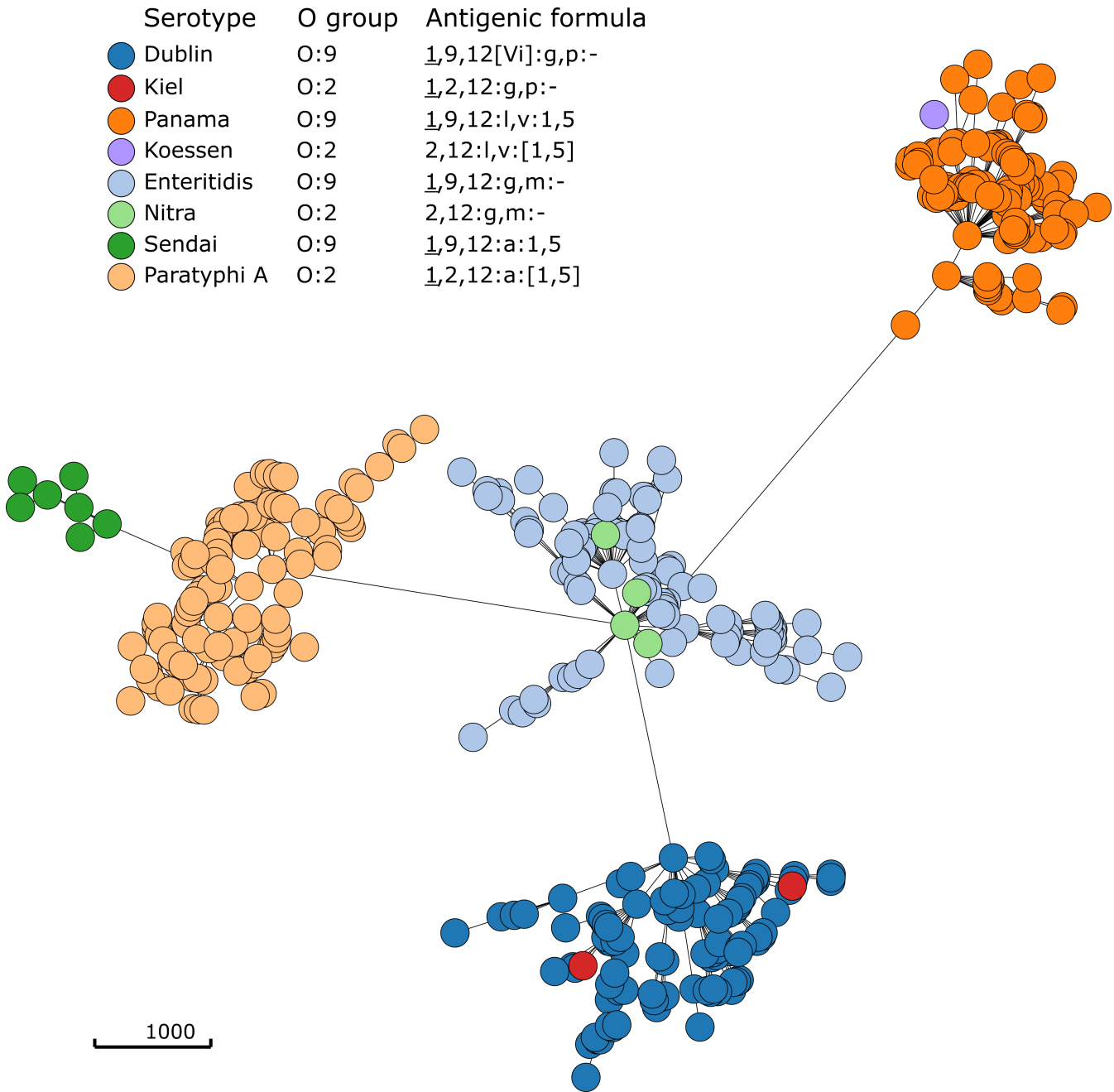
### Some serotypes are genetic variants of more common serotypes

Phylogenetic analysis of serogroups O:2 and O:9 serotypes that share the same H antigens revealed a close genetic relationship between serotypes that differ only by O group (Fig. 2). The four variants (Paratyphi A, Nitra, Kiel, and Koessen) belong to serogroup O:2 and are closely related to serogroup O:9 serotypes (Sendai, Enteritidis, Dublin, and Panama, respectively). Closely related serotypes and variants may be differentiated using additional subtyping methods. Seven-gene MLST functionality was added to SeqSero2S to improve subtype information available through the SeqSero2S analysis and to facilitate the identification of possible polyphyletic groups within a serotype, some of which may be a result of the simplified interpretation of serotypes. Figure 3A shows an example of the application of 7-gene MLST data to improve serotype discrimination for serotype Gallinarum. Gallinarum and Enteritidis have the same antigenic formula using genetic methods, I 9:g,m:-. Gallinarum is nonmotile phenotypically and does not express the H: g,m allele. They can be differentiated by phenotypic or genetic tests. SeqSero2S uses *sdf* to identify many Enteritidis strains (23); strains that are I 9:g,m:- and *sdf* negative are indicated as “Gallinarum or Enteritidis.” In this example, the Gallinarum genome was output as serotype “Gallinarum or Enteritidis” because of the absence of *sdf*. Its sequence type was determined to be ST92, which is associated with historical Gallinarum strains and distinct from the major ST183 lineage among isolates predicted as “Gallinarum or Enteritidis” by SeqSero2 in Enterobase (Fig. 3B). Of note,



**FIG 1** Survey of serotypes from WGS data, 2018 to 2023. (A) A Venn diagram showing the distribution of surveyed serotypes among humans, major food animals, and other sources. (B) A Venn diagram showing the distribution of surveyed serotypes in the United States and other countries. (C) A histogram of the 56 most reported serotypes worldwide that accounted for 90% of surveyed isolates. A word cloud insert highlights these serotypes with font sizes proportional to serotype proportions in the surveyed pool. The data for this figure and percentages of surveyed genomes by serotype are recorded in Table S3.

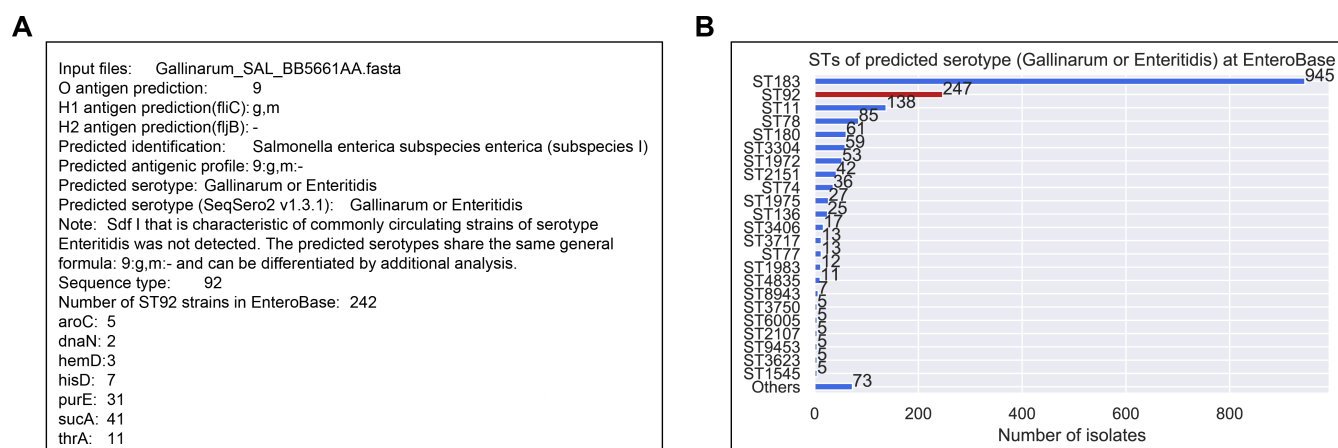
Gallinarum biovar Gallinarum isolates are ST78 whereas Gallinarum biovar Pullorum isolates are ST92.



**FIG 2** Phylogenetic relationship among representative genomes from serogroups O:2 and O:9. The minimum spanning tree is based on the EnteroBase cgMLST scheme. One hundred isolates of each serotype were randomly selected from EnteroBase. When 100 genomes were not available, all representative isolates were included.

### Summary of simplification

A total of 178 serotypes were provisionally removed (Table S1). Among these, 84 were excluded because their exact antigenic formula could not be determined, 66 because no alleles or probes were available, 10 because probes were unavailable for O:9,46,27, 14 because O:54 is encoded on a plasmid, and 4 because they could not be differentiated from other serotypes. A total of 58 serotypes were merged into other serotypes (Table S1). The antigenic formulae of 890 serotypes were simplified, but the simplifications did not result in their merging with other serotypes. Removed serotypes, merged serotypes, and simplified antigenic formulas are categorized in Table S1.



**FIG 3** Example of SeqSero2S output for two serotypes that have the same antigenic formula by genetic methods. (A) Textual output of serotype prediction including ST and corresponding 7-gene MLST loci. (B) Distribution of STs among all the “Gallinarum or Enteritidis” predictions by SeqSero2 at EnteroBase. The predicted ST of this query, ST92, is highlighted in red.

## DISCUSSION

SeqSero2 determines serotype from WGS data by targeting genetic determinants responsible for serotype, including subspecies markers and genes involved in the biosynthesis of O and H antigens. This approach maintains continuity with the phenotypic serotypes and achieves robust serotype prediction, including for serotypes that have not previously been characterized by genetic methods, and without consulting surrogate markers that are unrelated to serotype. The effectiveness of this approach has been validated by U.S. federal agencies (CDC, FDA, and USDA FSIS) for routine *Salmonella* surveillance, especially when serotype is the primary or only *Salmonella* subtype, e.g., surveillance systems NARMS and Laboratory-based Enteric Disease Surveillance. Serotype prediction by SeqSero2 is then followed by other WGS-based subtyping methods, such as genome-wide MLST and SNP analyses when desired.

We developed SeqSero2S to address gaps in our understanding of the genetic basis for some serotype antigens, thus limiting the representation of some serotypes and serotype antigens in a DNA-based scheme. The key aspects of the simplification include (i) disregard secondary O antigens that do not affect serotype assignment; (ii) disregard serotypes that are known to be minor genetic variants of more common serotypes and combine with the parent serotype, e.g., serogroup O:2 except serotype Paratyphi A; (iii) disregard secondary O antigens whose genetic basis is unknown; (iv) disregard antigens and antigenic types that are indicated as variables in the WKL Scheme to be consistent with predominant type for that particular serotype, and (v) disregard antigenic types for which an allele is unavailable or actual antigenic type is unknown due to variable epitopes in the antigenic formula. File S1 and Table S2 provide a detailed explanation of how SeqSero2S designates serotypes.

Our simplified interpretation of the WKL Scheme facilitates the adoption of universal typeability by a genetic approach to serotype determination by addressing multiple challenges in replicating a phenotypic scheme.

### Secondary O antigens in groups O:8 and O:6,14

In serogroup O:8, 39 pairs of serotypes differ by antigen O:6 according to the WKL Scheme. Mikoleit et al. characterized strains from five serotype pairs and found that all strains demonstrated variable expression of O:6, suggesting that O:8 serotypes differing by O:6 are not distinct serotypes (24). Also in serogroup O:8, secondary antigen O:20 is present only if the culture is lysogenized by an O:20 converting phage (phage conversion) (10). The genetic basis for O:20 is unknown. In serogroup O:6,14, two serotype-differentiating secondary O antigens O:24 and O:25 may be variably present



without relation to lysogenic conversion (10). Their genetic basis is unknown. In the simplified interpretation, serotypes that differed by O:6 and/or O:20 were merged, as were serotypes that differed by O:24 and O:25. When two serotypes are merged, the name of the serotype that was reported first is retained as detailed in Section V, File S1.

### **Antigenic types that are phylogenetically nested within more common types (e.g., serogroup O:2, H antigens of the G complex)**

Serogroup O:2 serotypes appear to be closely related to O:9 serotypes with the same H antigens (Fig. 2). The genetic basis for serogroup O:2 in serotype Paratyphi A, a frame-shift mutation in an otherwise serogroup O:9 *rfb* region, has been described (25). The genetic basis for the other serotypes has not been described but may be due to similar mutations that result from immune pressure or genetic drift. Similarly, several antigenic types within the G complex are known to be minor genetic variants of more common types, e.g., H:g,m,q, H:g,q, H:g,z63, and H:g,z85 are minor variants of H:g,m in what is otherwise a serotype Enteritidis strain (26). Disregarding these antigenic types helps improve the clarity of *Salmonella* surveillance data by avoiding the confusion caused by closely related serotypes such as Enteritidis and Nitra.

### **Antigens and epitopes that are indicated as variables in the WKL Scheme**

Some antigens in the WKL Scheme are indicated in square brackets or parentheses, indicating that they are variably present or may be weakly expressed, respectively. Such variability is difficult to represent in a genetic scheme. Furthermore, currently circulating strains typically do not show this variability, making the inclusion of variable epitopes in antigenic formulas confusing in genetic serotype prediction. For the simplification, we disregarded this variability when a definitive antigenic type was known for a particular serotype, e.g., the antigenic formula for serotype Montevideo is I 6,7,14,[54]:g,m,[p],s:[1,2,7] but currently circulating strains are predominantly I 6,7:g,m,s:-.

### **Lack of alleles for some antigenic types**

Alleles are not available for some antigenic types described in the WKL Scheme, e.g., H:z42, H:g,s,q, H:1,2,5. Also, some antigenic types could not be verified using well-characterized strains. In those instances, the serotype is indicated as “provisionally removed from the scheme” in Table S2 to indicate that SeqSero2S cannot identify that serotype.

### **Monophasic variants**

Like SeqSero2, SeqSero2S correctly identifies monophasic variants when the *fljB* gene is deleted. For instance, a common monophasic variant I 4,[5],12:i:-, many of which are variants of Typhimurium, is not recognized in the WKL Scheme but is reported as I 4:i- (after the O group simplification) with a note indicating it is a potential variant of Typhimurium. Similarly, I 4:b:- is reported and noted as a potential monophasic variant of Paratyphi B. When a monophasic variant results from a lack of expression or a point mutation, the gene is still present and both SeqSero2 and SeqSero2S predict these strains as the parent serotype. The added MLST functionality of SeqSero2S aids users in interpreting monophasic variants.

To circumvent the incomplete understanding of genetic determinants for some serotypes and to acknowledge that some serotypes are polyphyletic (27), some authors have proposed replacing (26), deducing (27), or affirming (16) serotypes using genetic markers unrelated to serotype. A common surrogate is sequence type (ST) determined by 7-gene MLST (26, 27). Using the correlation between some STs and serotypes, MLST alone was evaluated for serotype prediction for *Salmonella* isolates from routine public health surveillance in the United Kingdom (27). While this method avoided the perceived complexity associated with genetic determinants of serotype and was accurate for commonly circulating strains, misassignment of serotypes occurred when two different serotypes belonged to the same ST (27). In fact, 7-gene MLST was used to show

many serotypes were distributed across multiple unrelated phylogenetic clusters (26). Polyphyletic serotypes confound the recognition of evolutionary relatedness by serotype and serve as a main argument to replace serotype with MLST (26). In addition to the polyphyly, another confounding factor of surrogating serotype determination with ST is that some closely related serotypes can share the same ST, such as Paratyphi A and Sendai (ST85). While SeqSero2S incorporates the 7-gene MLST scheme as a complementary analysis, its serotype prediction operates independently without consulting ST. This contrasts with SISTR, which uses core genome MLST to supplement serotype prediction (16).

Genetic determination of serotypes implemented in SeqSero and SeqSero2 is based on the genes responsible for serotypes to replicate traditional serotypes as closely as possible (15, 17). Serotyping is by definition a phenotypic subtyping method; it comes as no surprise that it does not reflect genetic relationships. Modification of the current phenotype-based scheme to try to reflect phylogeny may add to the confusion that already exists regarding *Salmonella* serotypes. Serotyping and MLST, for example, are distinct and complementary subtyping methods. Both can be automatically determined from WGS data, and observations and decisions can be made based on the availability of two complementary subtypes.

We understand the appeal of defining serotypes using additional markers that reflect genetic relatedness; this simplified scheme furthers the goal of concordance of serotypes with natural population structure without the need for additional markers. Furthermore, the WKL Scheme is internationally recognized and accepted. The logistics of developing, obtaining international acceptance, and maintaining a new scheme that includes surrogate genetic markers seems unfeasible. Furthermore, maintaining a concordance between phenotypic and genetic serotypes provides continuity with laboratories that do not have WGS capacity and perform phenotypic serotyping.

The proposed simplification addresses multiple needs in the era of genomic prediction of serotypes. First, it addresses the fact that some serotypes in the WKL Scheme cannot be fully or accurately described using genetic methods. Second, the proposed simplification can be implemented under the current version of the WKL Scheme without publishing a new version of the scheme. Historical data can easily be updated to the simplified scheme if desired. Third, the proposed simplified scheme remains compatible with phenotypic serotyping, allowing continued determination of serotypes by phenotypic methods where needed. Finally, like the phenotypic scheme, this genetic scheme can be updated as recognized serotypes are further characterized or new antigenic types are discovered. This genetic scheme, along with the WKL Scheme, will continue to be improved as new studies enhance our understanding of antigenic diversity and the biological basis for serotype antigens.

A potential drawback of this approach is that merged serotypes may be common in some parts of the world or they may become important in the future, warranting their specific recognition. In the survey of worldwide *Salmonella* genomic surveillance between 2018 and 2023 (Fig. 1), 90% of genomes belonged to 56 serotypes, none of which are affected by the simplification. The simplified scheme can be updated as needed if new data warrants new or modified serotype definitions. Another potential drawback is the removal of secondary O antigens which may result in the merger of distinct genetic lineages into a single “serotype.” The current scheme already describes many serotypes that contain multiple genetic lineages (26). For example, serotype Kentucky contains two distinct lineages; ST152 is commonly found in poultry in the United States and an MDR lineage of ST198 has been associated with human infections in other parts of the world (28, 29). The simplified scheme may add a few more serotypes to that list. Also, when minor variants of a particular antigenic type are merged with the more common type, possible outbreaks might be missed in surveillance data unless that strain has other distinguishing characteristics such as 7-gene MLST or cgMLST type. We added 7-gene MLST analysis to SeqSero2S, including presenting prevalent STs of the predicted serotype, to facilitate the identification of polyphyletic serotypes. Finally, the



removal of secondary antigens from the scheme does not preclude continued characterization of them, if desired. The difference is they are no longer required to assign a “serotype” in SeqSero2S. For example, O:5 is an ancillary antigen found in some group O:4 serotypes; *oafA* has been reported to be responsible for this epitope. O:5 has been used to differentiate within serotype Typhimurium, O:5- variants were described as “var. Copenhagen.” Some, but not all, O:5- variants have an *oafA* allele that contains a 7 bp deletion that inactivates the gene (30). These variants are identified as we previously implemented (17).

## MATERIALS AND METHODS

### Survey of circulating serotypes using genomic surveillance data

We surveyed “computed types” (SeqSero2 predicted serotypes) of 32,499 genomes from *Salmonella* isolates deposited at NCBI Pathogen Detection ([https://www.ncbi.nlm.nih.gov/pathogens/isolates/#taxgroup\\_name:%22Salmonella%20enterica%22](https://www.ncbi.nlm.nih.gov/pathogens/isolates/#taxgroup_name:%22Salmonella%20enterica%22)) from 2018 to 2023, excluding 972 genomes with no or incomplete antigen predictions. These genomes were from major U.S. contributors including PulseNet USA (<https://www.cdc.gov/pulsenet/index.html>, human isolates from sporadic and outbreak cases), the National Antimicrobial Resistance Monitoring System for Enteric Bacteria (<https://www.cdc.gov/narms/index.html>; isolates from humans, retail meats, and food animals), and GenomeTrakr (<https://www.fda.gov/food/whole-genome-sequencing-wgs-program/genometrakr-network>, isolates from non-human domestic and imported sources [food, food facilities, animals, farms, water, etc.]). Only genomes whose location of isolation was the United States were included, although some of these isolates might have been sampled from imported commodities or patients who contracted *Salmonella* abroad. Similarly, we surveyed SeqSero2-predicted serotypes of 58,364 isolates deposited at Enterobase (<https://enterobase.warwick.ac.uk/species/index/senterica>) from 2018 to 2023 that did not overlap with the 32,499 genomes from NCBI and came from 108 countries. Similarly, 29,526 Enterobase genomes with no or incomplete antigen predictions by the Enterobase implementation of SeqSero2 (genome assembly mode, version <v1.2.1) were excluded. Finally, we excluded 201 serotypes that were represented by only one isolate in NCBI and Enterobase combined.

### Simplified interpretation of the WKL Scheme

A detailed description of the simplified interpretation of the serotype is presented in File S1. When the simplified interpretation resulted in merging two or more serotypes, we retained the name of the earliest described serotype (Table S1). Simplification rationales are detailed case-by-case in the Discussion.

### SeqSero2S

We implemented the simplification plan in a bioinformatic pipeline termed SeqSero2S. Antigenic types recognized by SeqSero2S are listed in Table S2. We added the 7-gene MLST (31) to SeqSero2S as a separate, complementary subtyping method. The draft or complete assembly of a query genome was scanned against the PubMLST scheme (32) using <https://github.com/tseemann/mlst> to determine its ST in addition to its serotype. ST data were analyzed in comparison to SeqSero2-predicted serotypes for over 400,000 *Salmonella* isolates deposited in Enterobase (33) as of May 2023.

## ACKNOWLEDGMENTS

We thank Sunee Himathongkham for coordinating the multi-agency consortium on the simplified interpretation of the *Salmonella* serotypes. We authors also thank Glenn

Tillman with USDA FSIS and Satheesh Nair of The UK Health Security Agency for their review and input.

The opinions expressed in this article are those of the authors and not necessarily those of the FDA, FSIS, or the CDC. Use of trade names is for identification only and does not imply endorsement by the FDA, the Public Health Service, the U.S. Department of Health and Human Services, or the U.S. Department of Agriculture.

## AUTHOR AFFILIATIONS

<sup>1</sup>Center for Food Safety, University of Georgia, Griffin, Georgia, USA

<sup>2</sup>School of Biomedical and Pharmaceutical Sciences, Guangdong University of Technology, Guangzhou, China

<sup>3</sup>Soochow University Cancer Institute, Soochow University, Suzhou, China

<sup>4</sup>Office of Regulatory Science, Office of Regulatory Affairs, U.S. Food and Drug Administration, Rockville, Maryland, USA

<sup>5</sup>Center for Food Safety and Applied Nutrition, U.S. Food and Drug Administration, College Park, Maryland, USA

<sup>6</sup>Center for Veterinary Medicine, U.S. Food and Drug Administration, Laurel, Maryland, USA

<sup>7</sup>Division of Foodborne, Waterborne and Environmental Diseases, Centers for Disease Control and Prevention, Atlanta, Georgia, USA

<sup>8</sup>Institut Pasteur, Université Paris Cité, Paris, France

## AUTHOR ORCIDs

Xiangyu Deng  <http://orcid.org/0000-0002-7251-2529>

Tongzhou Xu  <http://orcid.org/0000-0002-3829-8688>

Zhemín Zhou  <http://orcid.org/0000-0001-9783-0366>

Ruth Timme  <http://orcid.org/0000-0002-9705-5897>

Patricia I. Fields  <http://orcid.org/0000-0003-1199-2650>

## AUTHOR CONTRIBUTIONS

Xiangyu Deng, Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Writing – original draft, Writing – review and editing | Shaoting Li, Formal analysis, Software, Supervision | Tongzhou Xu, Formal analysis, Investigation, Software, Visualization | Zhemín Zhou, Resources | Michelle M. Moore, Investigation, Writing – review and editing | Ruth Timme, Investigation, Writing – review and editing | Shaohua Zhao, Investigation, Writing – review and editing | Charlotte Lane, Investigation, Writing – review and editing | Blake A. Dinsmore, Investigation, Writing – review and editing | François-Xavier Weill, Investigation, Writing – review and editing | Patricia I. Fields, Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Writing – original draft, Writing – review and editing

## DIRECT CONTRIBUTION

This article is a direct contribution from Xiangyu Deng, a member of the *Applied and Environmental Microbiology* Editorial Board, who arranged for and secured reviews by Ruiting Lan, University of New South Wales, and Rachel A. Cheng, Virginia Polytechnic Institute and State University.

## ADDITIONAL FILES

The following material is available [online](#).

## Supplemental Material

**File S1 (AEM02600-24-s0001.docx).** Conventions for simplified serotype designations using genetic methods implemented in SeqSero2S (SS2S).

**Table S1 (AEM02600-24-s0002.xlsx).** Simplified scheme.

**Table S2 (AEM02600-24-s0003.xlsx).** Antigenic types detected by SeqSero2.

**Table S3 (AEM02600-24-s0004.xlsx).** Serotypes surveyed from NCBI Pathogen Detection and Enterobase.

## REFERENCES

- Tack DM, Ray L, Griffin PM, Cieslak PR, Dunn J, Rissman T, Jervis R, Lathrop S, Muse A, Duwell M, Smith K, Tobin-D'Angelo M, Vugia DJ, Zablotsky Kufel J, Wolpert BJ, Tauxe R, Payne DC. 2020. Preliminary incidence and trends of infections with pathogens transmitted commonly through food - foodborne diseases active surveillance network, 10 U.S. sites, 2016-2019. *MMWR Morb Mortal Wkly Rep* 69:509-514. <https://doi.org/10.15585/mmwr.mm6917a1>
- Shah HJ, Jervis RH, Wymore K, Rissman T, LaClair B, Boyle MM, Smith K, Lathrop S, McGuire S, Trevejo R, McMillian M, Harris S, Zablotsky Kufel J, Houck K, Lau CE, Devine CJ, Boxrud D, Weller DL. 2024. Reported incidence of infections caused by pathogens transmitted commonly through food: impact of increased use of culture-independent diagnostic tests - foodborne diseases active surveillance network, 1996-2023. *MMWR Morb Mortal Wkly Rep* 73:584-593. <https://doi.org/10.15585/mmwr.mm7326a1>
- Delahoy MJ, Shah HJ, Weller DL, Ray LC, Smith K, McGuire S, Trevejo RT, Scallan Walter E, Wymore K, Rissman T, McMillian M, Lathrop S, LaClair B, Boyle MM, Harris S, Zablotsky-Kufel J, Houck K, Devine CJ, Lau CE, Tauxe RV, Bruce BB, Griffin PM, Payne DC. 2023. Preliminary incidence and trends of infections caused by pathogens transmitted commonly through food - foodborne diseases active surveillance network, 10 U.S. sites, 2022. *MMWR Morb Mortal Wkly Rep* 72:701-706. <https://doi.org/10.15585/mmwr.mm7226a1>
- Salmonella* Subcommittee of the Nomenclature Committee of the International Society for Microbiology. 1934. The genus *Salmonella* lignieres, 1900. *J Hyg (Lond)* 34:333-350. <https://doi.org/10.1017/s0022172400034677>
- CDC. 2013. An atlas of *Salmonella* in the United States, 1968-2011. Available from: <https://www.cdc.gov/salmonella/pdf/salmonella-atlas-508c.pdf>
- Coburn B, Grassl GA, Finlay BB. 2007. *Salmonella*, the host and disease: a brief review. *Immunol Cell Biol* 85:112-118. <https://doi.org/10.1038/sj.icb.7100007>
- Bäumler AJ, Tsois RM, Ficht TA, Adams LG. 1998. Evolution of host adaptation in *Salmonella enterica*. *Infect Immun* 66:4579-4587. <https://doi.org/10.1128/IAI.66.10.4579-4587.1998>
- Galanis E, Lo Fo Wong DMA, Patrick ME, Binsztein N, Cieslik A, Chalermchikit T, Aidara-Kane A, Ellis A, Angulo FJ, Wegener HC, World Health Organization Global Salm-Surv. 2006. Web-based surveillance and global *Salmonella* distribution, 2000-2002. *Emerg Infect Dis* 12:381-388. <https://doi.org/10.3201/eid1205.050854>
- Jackson BR, Griffin PM, Cole D, Walsh KA, Chai SJ. 2013. Outbreak-associated *Salmonella enterica* serotypes and food Commodities, United States, 1998-2008. *Emerg Infect Dis* 19:1239-1244. <https://doi.org/10.3201/eid1908.121511>
- Patrick G, François-Xavier W. 2007. Antigenic formulae of the *Salmonella* serovars. 9th ed. WHO Collaborating Centre for Reference and Research on *Salmonella*.
- Guibourdenche M, Roggentin P, Mikoleit M, Fields PI, Bockemühl J, Grimont PAD, Weill F-X. 2010. Supplement 2003-2007 (no. 47) to the White-Kauffmann-Le Minor scheme. *Res Microbiol* 161:26-29. <https://doi.org/10.1016/j.resmic.2009.10.002>
- Issenhuht-Jeanjean S, Roggentin P, Mikoleit M, Guibourdenche M, de Pinna E, Nair S, Fields PI, Weill F-X. 2014. Supplement 2008-2010 (no. 48) to the White-Kauffmann-Le Minor scheme. *Res Microbiol* 165:526-530. <https://doi.org/10.1016/j.resmic.2014.07.004>
- Fitzgerald C, Collins M, van Duyn S, Mikoleit M, Brown T, Fields P. 2007. Multiplex, bead-based suspension array for molecular determination of common *Salmonella* serogroups. *J Clin Microbiol* 45:3323-3334. <https://doi.org/10.1128/JCM.00025-07>
- McQuiston JR, Waters RJ, Dinsmore BA, Mikoleit ML, Fields PI. 2011. Molecular determination of H antigens of *Salmonella* by use of a microsphere-based liquid array. *J Clin Microbiol* 49:565-573. <https://doi.org/10.1128/JCM.01323-10>
- Zhang S, Yin Y, Jones MB, Zhang Z, Deatherage Kaiser BL, Dinsmore BA, Fitzgerald C, Fields PI, Deng X. 2015. *Salmonella* serotype determination utilizing high-throughput genome sequencing data. *J Clin Microbiol* 53:1685-1692. <https://doi.org/10.1128/JCM.00323-15>
- Yoshida CE, Kruczkiewicz P, Laing CR, Lingohr EJ, Gannon VPJ, Nash JHE, Taboada EN. 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. <https://doi.org/10.1371/journal.pone.0147101>
- Zhang S, den Bakker HC, Li S, Chen J, Dinsmore BA, Lane C, Lauer AC, Fields PI, Deng X. 2019. SeqSero2: rapid and improved *Salmonella* serotype determination using whole-genome sequencing data. *Appl Environ Microbiol* 85:e01746-19. <https://doi.org/10.1128/AEM.01746-19>
- Swaminathan B, Barrett TJ, Hunter SB, Tauxe RV, Force C. 2001. PulseNet: the molecular subtyping network for foodborne bacterial disease surveillance, United States. *Emerg Infect Dis* 7:382-389. <https://doi.org/10.3201/eid0703.010303>
- Allard MW, Strain E, Melka D, Bunning K, Musser SM, Brown EW, Timme R. 2016. Practical value of food pathogen traceability through building a whole-genome sequencing network and database. *J Clin Microbiol* 54:1975-1983. <https://doi.org/10.1128/JCM.00081-16>
- Jackson BR, Tarr C, Strain E, Jackson KA, Conrad A, Carleton H, Katz LS, Stroika S, Gould LH, Mody RK, et al. 2016. Implementation of nationwide real-time whole-genome sequencing to enhance listeriosis outbreak detection and investigation. *Clin Infect Dis* 63:380-386. <https://doi.org/10.1093/cid/ciw242>
- Katz LS, Griswold T, Williams-Newkirk AJ, Wagner D, Petkau A, Sieffert C, Van Domselaar G, Deng X, Carleton HA. 2017. A comparative analysis of the Lyve-SET phylogenomics pipeline for genomic epidemiology of foodborne pathogens. *Front Microbiol* 8:375. <https://doi.org/10.3389/fmicb.2017.00375>
- Tolar B, Joseph LA, Schroeder MN, Stroika S, Ribot EM, Hise KB, Gerner-Smidt P. 2019. An overview of PulseNet USA databases. *Foodborne Pathog Dis* 16:457-462. <https://doi.org/10.1089/fpd.2019.2637>
- Agron PG, Walker RL, Kinde H, Sawyer SJ, Hayes DC, Wollard J, Andersen GL. 2001. Identification by subtractive hybridization of sequences specific for *Salmonella enterica* serovar enteritidis. *Appl Environ Microbiol* 67:4984-4991. <https://doi.org/10.1128/AEM.67.11.4984-4991.2001>
- Mikoleit M, Van Duyn MS, Halpin J, McGlinchey B, Fields PI. 2012. Variable expression of O:6 in *Salmonella* group C2. *J Clin Microbiol* 50:4098-4099. <https://doi.org/10.1128/JCM.01676-12>
- Verma NK, Quigley NB, Reeves PR. 1988. O-antigen variation in *Salmonella* spp.: rfb gene clusters of three strains. *J Bacteriol* 170:103-107. <https://doi.org/10.1128/jb.170.1.103-107.1988>
- Achtman M, Wain J, Weill F-X, Nair S, Zhou Z, Sangal V, Krauland MG, Hale JL, Harbottle H, Uesbeck A, Dougan G, Harrison LH, Brisse S, S. Enterica MLST Study Group. 2012. Multilocus sequence typing as a replacement for serotyping in *Salmonella enterica*. *PLoS Pathog* 8:e1002776. <https://doi.org/10.1371/journal.ppat.1002776>
- Ashton PM, Nair S, Peters TM, Bale JA, Powell DG, Painset A, Tewolde R, Schaefer U, Jenkins C, Dallman TJ, de Pinna EM, Grant KA, Salmonella Whole Genome Sequencing Implementation Group. 2016. Identification of *Salmonella* for public health surveillance using whole genome sequencing. *PeerJ* 4:e1752. <https://doi.org/10.7717/peerj.1752>
- Hawkey J, Le Hello S, Doublet B, Granier SA, Hendriksen RS, Fricke WF, Ceysens PJ, Gomart C, Billman-Jacobe H, Holt KE, Weill FX. 2019. Global

- phylogenomics of multidrug-resistant *Salmonella enterica* serotype Kentucky ST198. *Microb Genom* 5:e000269. <https://doi.org/10.1099/mgen.0.000269>
29. Haley BJ, Kim SW, Pettengill J, Luo Y, Karns JS, Van Kessel JAS. 2016. Genomic and evolutionary analysis of two *Salmonella enterica* serovar Kentucky sequence types isolated from bovine and poultry sources in North America. *PLoS One* 11:e0161225. <https://doi.org/10.1371/journal.pone.0161225>
30. Hauser E, Junker E, Helmuth R, Malorny B. 2011. Different mutations in the *oafA* gene lead to loss of O5-antigen expression in *Salmonella enterica* serovar Typhimurium. *J Appl Microbiol* 110:248–253. <https://doi.org/10.1111/j.1365-2672.2010.04877.x>
31. Maiden MC, Bygraves JA, Feil E, Morelli G, Russell JE, Urwin R, Zhang Q, Zhou J, Zurth K, Caugant DA, Feavers IM, Achtman M, Spratt BG. 1998. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc Natl Acad Sci U S A* 95:3140–3145. <https://doi.org/10.1073/pnas.95.6.3140>
32. Jolley KA, Bray JE, Maiden MCJ. 2018. Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications. *Wellcome Open Res* 3:124. <https://doi.org/10.12688/wellcomeopenres.14826.1>
33. Zhou Z, Alikhan NF, Mohamed K, Fan Y, Agama Study G, Achtman M. 2020. The Enterobase user's guide, with case studies on *Salmonella* transmissions, *Yersinia pestis* phylogeny, and *Escherichia* core genomic diversity. *Genome Res* 30:138–152. <https://doi.org/10.1101/gr.251678.119>