

Development of a robust FT-IR typing system for *Salmonella enterica*, enhancing performance through hierarchical classification

Diego Fredes-García,¹ Javiera Jiménez-Rodríguez,² Alejandro Piña-Iturbe,¹ Pablo Caballero-Díaz,¹ Tamara González-Villarroel,³ Fernando Dueñas,⁴ Aniela Wozniak,² Aiko D. Adell,⁵ Andrea I. Moreno-Switt,¹ Patricia García²

AUTHOR AFFILIATIONS See affiliation list on p. 12.

ABSTRACT *Salmonella enterica* is a leading cause of foodborne illnesses globally, with significant mortality rates, especially among vulnerable populations. Traditional serotyping methods for *Salmonella* are accurate but expensive, resource-intensive, and time-consuming, necessitating faster and more reliable alternatives. This study evaluates the IR Biotyper, a Fourier-transform infrared spectroscopy system, in differentiating *Salmonella* serovars. We assessed 458 isolates of nine *Salmonella* serovars (Infantis, Enteritidis, Typhimurium, I,4,[5],12:i:-, Montevideo, Agona, Thompson, Panama, and Abony) from diverse sources. The IR Biotyper was used to acquire spectra from these isolates. Machine learning algorithms, including support vector machines, were trained to classify the isolates. The accuracy of classifiers was validated using a validation set to determine sensitivity, specificity, positive predictive value, and negative predictive value. Initial classifiers showed high accuracy for Abony, Agona, Enteritidis, and Infantis serovars, with sensitivities close to 100%. However, classifiers for *S. Typhimurium*, *S. Panama*, and *S. Montevideo* exhibited lower performance. Implementing a hierarchical classification system enhanced the accuracy of serogroup O:4 serovars, demonstrating that this approach offers a robust framework for *Salmonella* serovar identification. The hierarchical system enables progressive refinement of classification, minimizing misclassifications by focusing on serogroup-specific features, making it adaptable to complex data sets and diverse serovars. The IR Biotyper demonstrates high potential for rapid and accurate *Salmonella* serovar identification. This study supports its implementation as a cost-effective, high-throughput tool for pathogen typing, enhancing real-time epidemiological surveillance, and guiding treatment strategies for salmonellosis. This method establishes a robust and scalable framework for advancing *Salmonella* serotyping practices across clinical, industrial, and public health domains by leveraging hierarchical classification.

IMPORTANCE Early and accurate identification of *Salmonella* serovars is extremely important for epidemiological surveillance, public health, and food safety. Traditional serotyping is very successful but is laborious and costly. In this study, we demonstrate the promise of Fourier-transform infrared spectroscopy together with machine learning as a means for *Salmonella* serotyping. Using hierarchical classification, we attain optimal serovar identification accuracy, particularly for challenging-to-type serogroups. Our findings recognize the IR Biotyper as a high-throughput, scalable pathogen typing solution that offers real-time data that can enable enhanced outbreak response and prevention of foodborne disease. The approach bridges the gap between traditional microbiological practice and sophisticated analytical technology, the path to more effective, cost-saving interventions in the clinical, industrial, and regulatory settings.

Editor Cristina Solano, Navarrabiomed-Universidad Pública de Navarra (UPNA)-Complejo Hospitalario de Navarra (CHN), IdiSNA, Pamplona, Navarra, Spain

Address correspondence to Patricia García, pgarcia@uc.cl.

The authors declare no conflict of interest.

Received 15 January 2025

Accepted 22 March 2025

Published 27 May 2025

Copyright © 2025 Fredes-García 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/).

Application of these technologies can significantly improve *Salmonella* surveillance-control and Public Health outcomes.

KEYWORDS *Salmonella*, serotyping, FT-IR, rapid diagnostic, salmonellosis

Salmonella enterica is one of the major causes of foodborne disease, causing more than 420,000 deaths worldwide, in 2019 only (1). While salmonellosis commonly presents as self-limited gastroenteritis, children under 5 years, adults over 65 years, and immunocompromised people are at a higher risk of developing a life-threatening severe invasive form of the disease which requires the use of third-generation cephalosporins, fluoroquinolones, or azithromycin as first-line antibiotics (2). The *Salmonella* genus has two species: *S. enterica* and *Salmonella bongori*. Within *S. enterica*, there are six subspecies; however, most human infection is caused by *S. enterica* subsp. *enterica* (3). Serological typing of the lipopolysaccharide (O) antigen results in more than 40 *Salmonella* serogroups, which, when phase 1 (H1) and phase 2 (H2) flagellar antigens are included, classify *Salmonella* in more than 2,600 serovars, according to the White-Kauffmann-Le Minor scheme (4, 5). Nevertheless, most human cases are caused by only a few serovars, such as *S. Enteritidis* and *S. Typhimurium* in Europe and the United States, where they are responsible for 30%-65% of *Salmonella* infections, for example (6, 7). Moreover, different lineages of these and other less frequent/emerging serovars (e.g., Infantis, Newport, Derby, and Heidelberg, among others) are constantly causing outbreaks of foodborne disease and have acquired antimicrobial resistance determinants against third-generation cephalosporins and fluoroquinolones, limiting the treatment options in cases of severe infection (6–12). Therefore, fast and reliable methods for identifying *Salmonella* serovars are required for performing real-time surveillance of high-risk/high-burden serovars within a One Health framework, that is, in humans, animals, and the environment, as control of human infections caused by this pathogen is influenced by its presence in animals and the environment (13).

Currently, the gold standard for *Salmonella* serovar identification is phenotypic serotyping using the White-Kauffmann-Le Minor scheme and antisera targeting the O, H1, and H2 antigens. However, this method requires more than 150 specific antisera and well-trained experienced personnel (14). Moreover, if only one H antigen is detected, additional steps are performed to confirm whether the tested strain is monophasic (has only one H antigen) or not. Therefore, traditional serotyping is usually carried out in reference laboratories and takes several days to produce results (14, 15). As alternatives to traditional serotyping, molecular methods based on detecting serovar-specific genetic markers are also available and can produce results in a reduced time (16). However, these methods require expensive reactions, additional steps of DNA extraction, and/or purification besides bacterial isolation and are focused on a limited number of relevant/common serovars (16, 17). Whole-genome sequencing (WGS) coupled with *in silico* serovar prediction has become a powerful tool for *Salmonella* serotyping, as it provides highly accurate typing results and supplies the raw material (the genome sequence) for further in-depth characterization of the pathogen (16, 18). Nevertheless, there are multiple barriers to the routine implementation of WGS in the clinical setting, especially for low-middle-income countries, including high setup and running costs, a lack of studies addressing the cost-effectiveness of WGS implementation, and logistic challenges about genomic data storage, analysis, and reporting (19). Cost-effective, high-throughput serotyping alternatives that can produce results in less than 24 hours are desirable.

The Fourier-transform infrared (FT-IR) spectroscopy is a vibrational spectroscopic technique that can be applied to assess the chemical composition of a sample (20). The different vibrational bond energies between the atoms in the sample molecules absorb IR radiation at different wavelengths, producing a spectrum characteristic of their composition (21). With multiple biological/biomedical applications, such as cell imaging, cancer detection, food safety, toxicology, microbiology, and others (21), FT-IR

spectroscopy is considered a versatile, fast, and low-cost technique (22). For bacterial typing, the spectra obtained by FT-IR spectroscopy can work as fingerprints that allow differentiation of bacteria at the species, subspecies, and even serogroup/serotype level (23–25), and its application for strain clonality studies, outbreak investigation, and bacterial identification is being studied, as it can provide results in less than 4 hours at a fraction of the cost of higher resolution techniques such as WGS (15, 26–29). The IR Biotyper (Bruker Daltonics GmbH & Co. KG, Germany) is an IR spectrometer coupled to a microplate compartment, which, together with the OPUS software, allows the acquisition and analysis of FT-IR spectra for multiple bacterial samples (30). The IR Biotyper can be trained with different bacteria with the aid of machine-learning algorithms to enable the typing of unknown samples. Consequently, the implementation of the IR Biotyper, as a low-cost/high-throughput tool for pathogen typing in the clinical microbiology laboratory, has been proposed (26, 27). Using this approach, the discrimination of clinically relevant *Salmonella* serogroups has been achieved with high accuracy (15). However, FT-IR-based typing of this pathogen at the serovar level is still challenging, especially when multiple serovars of the same serogroup are tested (15, 31). To overcome this difficulty, a step-wise approach for serovar classification has been proposed, in which the step of serovar classification should follow the first step of serogroup classification; however, this approach has not yet been evaluated (32). In this study, we tested the capabilities of the IR Biotyper using the step-wise approach for the classification of the major *Salmonella* serovars found in the clinical, environmental, and food/avian niches in Chile.

MATERIALS AND METHODS

Bacterial collection, culture media, and conditions

A total of 458 well-characterized isolates of *S. enterica* (corresponding to nine serovars) corresponding to *Salmonella* Infantis ($n = 163$), *Salmonella* Enteritidis ($n = 117$), *Salmonella* Typhimurium ($n = 51$), *Salmonella* I,4,[5],12: i- ($n = 16$), *Salmonella* Montevideo ($n = 6$), *Salmonella* Agona ($n = 63$), *Salmonella* Thompson ($n = 14$), *Salmonella* Panama ($n = 16$), and *Salmonella* Abony ($n = 12$) were used in this study (Table S1).

The isolates were provided by different institutions and investigation centers and were obtained from different sources, such as environmental water ($n = 277$) (33), swine stool ($n = 1$) (34), an equine hospital facility ($n = 9$) (35), a backyard poultry farm ($n = 8$) (36), poultry meat ($n = 35$) (12), and clinical human samples ($n = 12$) (37) from patient consultant at Red de Salud UC-CHRISTUS ($n = 114$). Permission to use the human clinical isolates was granted by the Comité Ético Científico de Ciencias de la Salud UC (Protocol ID 240621037).

In summary, the isolation and serotyping protocol were conducted by processing samples through traditional microbiological techniques. Strain confirmation was conducted using molecular methods as outlined by Toro et al. (33). Serovar identification was performed through the standard Kaufmann-White-LeMinor agglutination system, complemented by molecular approaches (36, 38), as well as *in silico* analysis based on WGS. Table S1 provides detailed information on the collection year, source, and methodology for each isolate.

Sample preparation

Strains were stored in 12%–20% glycerol stocks at -80°C before they were retrieved on trypticase soy agar (TSA, Biomérieux, France) and passed at least two times before subculturing the samples in TSA; each incubation was performed at $37 \pm 2^{\circ}\text{C}$ for 24 hours. The sample preparation adhered to the manufacturer's guidelines. Briefly, a 1- μL overloaded loop of bacterial colonies was taken from the densely populated area of the culture and suspended on 50 μL of a 70% ethanol solution within an IR Biotyper suspension vial. After vortexing, 50 μL of deionized water was introduced, and the

mixture was thoroughly mixed by vortexing. Next, 15 μ L of the bacterial suspension was dispensed in three separate technical replicates onto the 96-spot silicon IR Biotyper target and left to dry for 15–20 minutes at $35 \pm 2^\circ\text{C}$. For all isolates categorized, measurements were taken in three or more distinct biological replicates to ensure a more equitable number of spectra for each serovar. As for the quality control, the Infrared Test Standards (IRTS 1 and IRTS 2) provided in the IR Biotyper kit were resuspended in 100- μ L deionized water, and 60 μ L of absolute ethanol was added and thoroughly mixed. A 10- μ L suspension was then deposited in duplicate onto the IR Biotyper target and allowed to dry following the same procedure as the samples.

Spectra acquisition and analysis

Acquisition of spectra was conducted in the UC-Christus Clinical Microbiology laboratory. This process involved transmission mode within the spectral range of 3,996–500 cm^{-1} (mid-IR), utilizing the IRBT spectrometer and OPUS software version 8.2.28 from Bruker Optics GmbH & Co. KG. Spectra processing and visualization was executed using the IR Biotyper Client Software from Bruker Daltonics, specifically version 4.0, with default settings recommended by the manufacturer. Before acquiring sample spectra, quality control measurements (IRTS 1 and IRTS 2) were conducted in each run. For each isolate, we obtained at least three spectrums, and to ensure quality, some spectra were reprocessed to achieve three correct measurements, obtaining a total of 5,428 spectra for the training and validation steps.

Machine learning and development of automated classifiers

The strategy of classification was divided into steps. The first step corresponds to the determination of the “O” antigen (O:4, O:7, and O:9) using the “*Salmonella* serogroup” classifier of the IR Biotyper Client Software from Bruker Daltonics (version 4.0), and the second step performs the serovar classification. Machine learning (ML) algorithm support vector machine (SVM) was utilized to construct the serovar IR-BT automated classifiers. The ML undergoes training with a set of well-characterized isolates to identify the specific characteristics of each serovar, and ML algorithm learns to recognize knowns, obtaining a real-time classification.

Accuracy was automatically calculated by the IR Biotyper software using a confusion matrix. The matrix determines the class recall as the percentage of correctly predicted spectra among all true spectra, considering uncertain spectra as errors. The class precision is calculated as the ratio of true positive cases to the sum of true positive and false positive cases, corresponding to the positive predictive value (PPV). Additionally, overall accuracy is computed as the proportion of all true positive cases relative to the total number of cases. Sensitivity is defined as the ratio of true positive cases to the sum of true positive and false negative cases, excluding uncertain cases. Specificity is calculated by dividing the number of true negative cases by the sum of true negative and false positive cases. Meanwhile, the negative predictive value (NPV) is determined as the ratio of true negative cases to the sum of true negative and false negative cases.

To better understand the clustering performance of the serovars, we performed a linear discriminant analysis (LDA) using IR Biotyper Client software version 4.0. The parameters of the classification algorithms, such as the number of principal components (PCs) and the C-value or cost value, which evaluate the penalty associated with misclassification errors made by the model during training, were optimized through a trial-and-error process for maximum accuracy. The classifier version that produced the best results was selected and further validated with an external testing set to ensure robustness and reliability (32). The steps described above were performed using the IR Biotyper software version 4.0. The variance of each LDA was calculated by the software.

For more detailed information about the equipment software, visit <https://www.bruker.com/en/applications/microbiology-and-diagnostics/food-beverage-microbiology/ir-biotyper-for-food-microbiology.html> (30).

TABLE 1 Confusion matrix for the validation of O-antigen^a

		Class predicted				Class recall
		O:4	O:7	O:9	Uncertain	
Actual class	Panama	0	0	180	0	100%
	Thompson	0	200	0	0	100%
	Typhimurium/1,4,[5],12: i:-	801	0	0	1	100%
	Abony	103	0	0	0	100%
	Agona	889	0	0	0	100%
	Enteritidis	0	0	1,528	0	100%
	Infantis	0	1,681	0	0	100%
	Montevideo	0	45	0	0	100%
	Class precision	100%	100%	100%		

^aAccuracy: 100% (5,427/5,428).

O-antigen classifier validation

To validate the O-antigen classifier integrated into the equipment, a comprehensive analysis was conducted using the total of isolates ($n = 458$) corresponding to the 5,428 spectra, which included representations from the serogroups O:4 (*S. Typhimurium*, *S. I,4,[5],12:i:-*, *S. Abony*, *S. Agona*), O:7 (*S. Infantis*, *S. Montevideo*, *S. Thompson*), and O:9 (*S. Enteritidis*, and *S. Panama*).

Serogroup O:4 training and validation

Serovars with more than five isolates per type were chosen. The proportion of training and validation depends on the heterogeneity within each group (O:4, O:7, O:9). The total isolates, together with their corresponding spectra, were divided into two randomized data sets, and the larger percentage was used for training and the remaining for validation.

For the first training set, we focused on serogroup O:4, which includes the serovars spectra *S. Typhimurium* ($n = 583$) and *S. I,4,[5],12: i:-* ($n = 219$), *S. Agona* ($n = 889$), and *S. Abony* ($n = 103$). Serovars *S. Typhimurium* and *S. I,4,[5],12: i:-* were used as a single target called *S. Typhimurium/Monophasic* ($n = 802$). Out of the total spectra obtained ($n = 1,794$), 80% ($n = 1,460$) were included in the training set (Table S1), these spectra belong to a subset of isolates ($n = 109$). The remaining 20% of the spectra ($n = 334$) were used for the validation set; these spectra belong to the remaining isolates ($n = 33$).

Serogroup O:7 training and validation

In the second round of training, we focused on serovars from serogroup O:7, including spectrums of the serovars *S. Infantis* ($n = 1,681$), *S. Thompson* ($n = 200$), and *S. Montevideo* ($n = 45$), generating a total of 1,926 new spectra. Of these new spectra, 58% ($n = 1,113$) were incorporated into the training set; these spectra belong to a subset of isolates ($n = 104$). The remaining 42% ($n = 813$) were added to the original validation data set; these spectra belong to the remaining isolates ($n = 79$).

Serogroup O:9 training and validation

In the third round of training, we focused on serovars from serogroup O:9, including *S. Enteritidis* ($n = 1,528$) and *S. Panama* ($n = 180$). Out of a total of 1,708 spectra, 80% ($n = 1,393$) were used for classifier development; these spectra belong to a subset of isolates ($n = 107$), while the remaining 20% ($n = 315$) were included in the validation set; these spectra belong to the remaining isolates ($n = 26$).

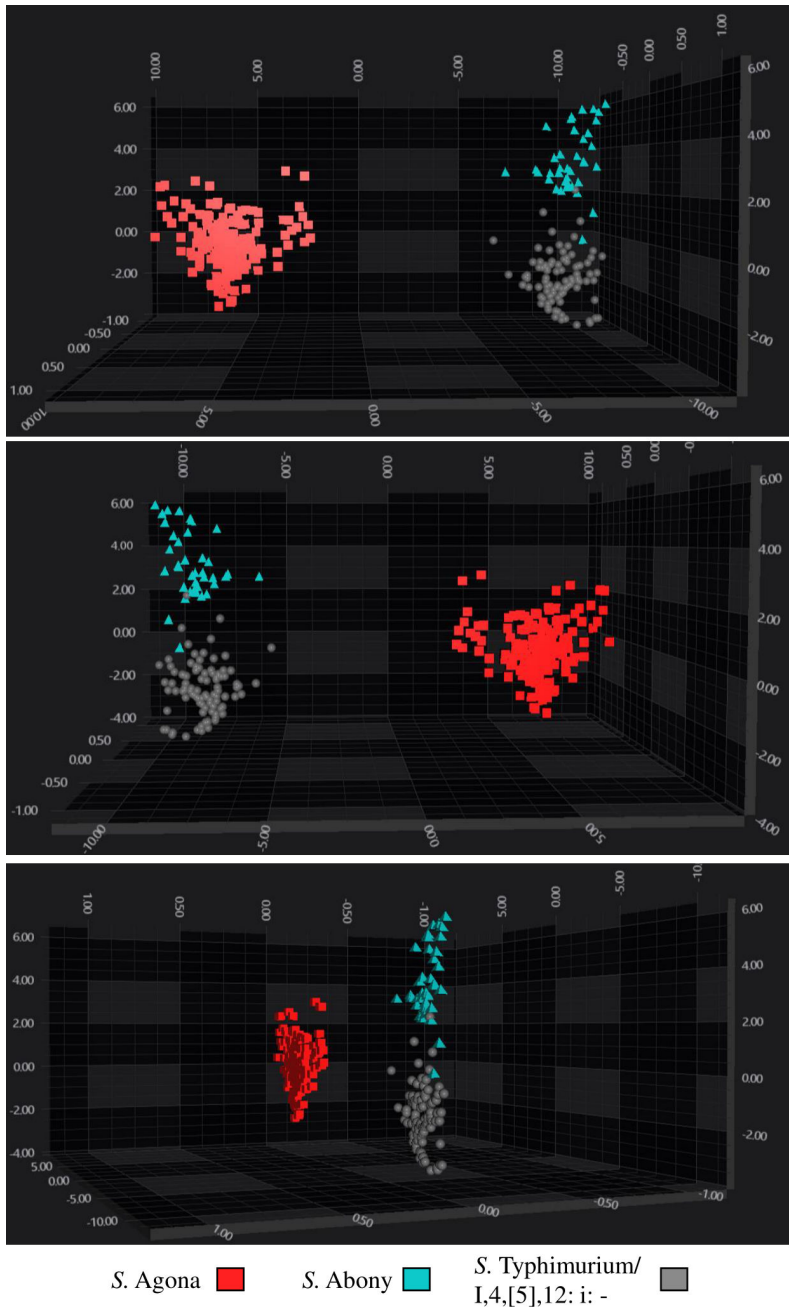


FIG 1 3D scatter plot of the third training performed, showing the distribution of serovars of *Salmonella* serogroup O:4 species in IR spectral space. A total of 30 PCs were used to create a LDA by serovar. The spectra are colored according to the serovars: red squares: *S. Agona*; cyan triangles: *S. Abony*; gray circles: *S. Typhimurium/I,4,[5],12:-*. Every symbol represents a different isolate.

RESULTS

O-antigen classifier validation

In the analysis of the total of 5,428 spectra, the O-antigen classifier validation achieved 100% accuracy, correctly predicting the serogroup of each one of the spectra; detailed accuracy values are provided in Table 1.

TABLE 2 Confusion matrix for the validation of O:4 classifier^a

		Class predicted				
		Typhimurium/ 1,4,[5],12: i:-	Abony	Agona	Uncertain	Class recall
Actual class	Typhimurium/ 1,4,[5],12: i:-	89	0	2	5	93% (89/96)
	Abony	0	38	0	0	100% (38/38)
	Agona	0	0	198	2	99% (198/200)
	Class precision	100%	100%	99%		

^aAccuracy: 97% (325/334).

Serogroup O:4 training and validation

The ML algorithm used to construct the initial training model was a Library for Support Vector Machines with a Radial Basis Function (LIB_SVM_RBF). This process involved the spectral range of 1,300–800 cm⁻¹ and was performed with PC30 and a C-value of 50 (Fig. 1). In total, 334 spectra were analyzed to perform the validation step. The automated classifiers for *S. Abony* showed the best performance (Table 2), achieving 100% sensitivity, specificity, PPV, and NPV, followed by the *S. Agona* and *S. Typhimurium*/1,4,[5], 12: i:- (Table 3). The LDA analysis showed that the isolates belonging to the *Agona* serovar exhibited a better clustering performance compared to the rest of the developed classifiers, followed by the serovar *Abony* and then *S. Typhimurium*/1,4,[5],12: i:-.

Serogroup O:7 training and validation

The ML algorithm developed for this training was LIB_SVM_RBF, which involved a spectral range of 1,300–800 cm⁻¹. SVM was performed with PC30 and a C-value of 50. Following validation, the *S. Infantis* classifier demonstrated the highest performance (Table 4), achieving 100% sensitivity, 94% specificity, 99% PPV, and 100% NPV. The *S. Thompson* classifier had the second-best performance, with 98% sensitivity, 100% specificity, 100% PPV, and 99% NPV. Finally, the *S. Montevideo* classifier achieved 67% sensitivity, 100% specificity, 100% PPV, and 99% NPV (Table 5). The LDA clustering analysis showed three well-separated groups of spectra: *Thompson* (red) exhibits more dispersion among the spectra; on the other hand, *Infantis* (grey) and *Montevideo* (cyan), both groups of spectra are more concentrated and exhibit well-characterized clusters (Fig. 2).

Serogroup O:9 training and validation

This training utilized 1,393 spectra, employing SVM across a spectral range of 1,300–800 cm⁻¹, with parameters set to PC30 and C-value 2. Following validation set analysis, the best-performing classifier was *S. Enteritidis* (Table 6), achieving 100% sensitivity, 80% specificity, 98% PPV, and 100% NPV and *S. Panama* with 80% sensitivity, 100% specificity, 100% PPV, and 98% NPV (Table 7). Analysis of LDA clustering showed a distinct separation of *S. Enteritidis* (grey) and *S. Panama* (red) (Fig. 3).

DISCUSSION

S. enterica is one of the most significant foodborne pathogens globally, posing substantial threats to both economic stability and public health. Identifying the specific serovar causing a clinical case is critical, as it allows for rapid and effective interventions, thereby mitigating outbreaks and preventing the further spread of infection. In this study, we propose a novel sub-typing method for *Salmonella* based on FT-IR, emphasizing

TABLE 3 Values of sensitivity, specificity, PPV, and NPV per serogroup O:4

	Sensitivity	Specificity	PPV	NPV
Typhimurium/ 1,4,[5],12: i:-	98%	100%	100%	99%
Abony	100%	100%	100%	100%
Agona	100%	98%	99%	100%

TABLE 4 Confusion matrix for the validation of O:7 classifier^a

		Class predicted				
		Thompson	Infantis	Montevideo	Uncertain	Class recall
Actual class	Thompson	103	2	0	0	98% (103/105)
	Infantis	0	691	0	2	99% (691/693)
	Montevideo	0	5	10	0	67% (10/15)
	Class precision	100%	99%	100%		

^aAccuracy: 99% (804/813).

its advantages over traditional or existing approaches. By leveraging discriminatory features identified during training, our classifiers enable the automatic categorization of unidentified samples using a marker model computed from a collection of training spectra. These automated classifiers are designed to seamlessly categorize isolates during spectrum measurement, with a particular focus on *S. enterica* serovars. To achieve our goal of developing a rapid and accurate *Salmonella* typing method, it is important to acknowledge the complexity involved. Differentiation among the approximately 2,600 serovars of *Salmonella* primarily relies on antigenic (phenotypic) distinctions.

As an initial step, the serogroup classifier was validated using the complete set of spectra, demonstrating high performance for each serogroup classifier (Table 1). Based on these results, we can infer that the integrated algorithm exhibits robust discriminatory capacity, effectively clustering spectra according to their serogroups. Similar findings were reported by Campos et al. (31), who evaluated the discriminatory power of FT-IR in distinguishing *Salmonella* serogroups and serovars. By developing algorithms to differentiate various serogroups and serovars within the same serogroup, their study found that the algorithm achieved high performance in discriminating among serogroups O:9, O:7, O:1, O:3, O:19, and O:4. Furthermore, the algorithm demonstrated high typification accuracy in distinguishing between serovars within the same serogroup, indicating its effectiveness in identifying and grouping spectrum features by serovar. It should be noted that different sources of isolates were used (animal, human, and environmental); however, no differences were evident based on the source of isolates. This could be attributed to the fact that most of the serovar analyzed are zoonotic, so it would be expected that there would be no differences between the different sources. Therefore, the possibility of overlap between serovars of the same serogroup could be attributed to structural similarities.

Based on the validation results for the integrated serogroup classifier and building upon previous studies (15, 31, 39), we propose the development of a subtyping algorithm to identify serovars from specific serogroups. This approach leverages the existing serogroup training incorporated into the IR Biotyper software developed by Cordovana et al. (15). This methodology uses a pre-established serogroup classification algorithm as the initial analytical step, followed by a serogroup-specific serovar classification algorithm to enhance typing performance by minimizing variability among trained features.

The best-performing classifiers validations for each serogroup were *S. Abony* (O:4), *S. Infantis* (O:7), and *S. Enteritidis* (O:9). The analysis of the O:4 classifier results reveals three distinct clusters, each representing a serovar: *S. Abony*, *S. Agona*, and *S. Typhimurium*/1,4,[5],12:i:-. Among them, the *S. Agona* cluster is the most clearly definierouped closely together and distinctly separated from the others. In contrast, the *S. Abony* and *S. Typhimurium*/1,4,[5],12:i:- clusters are positioned near each other, obtaining similar results in the training confusion matrix (data not shown). Although this suggests that

TABLE 5 Values of sensitivity, specificity, PPV, and NPV per serogroup O:7

	Sensitivity	Specificity	PPV	NPV
Thompson	98%	100%	100%	99%
Infantis	100%	94%	99%	100%
Montevideo	67%	100%	100%	99%

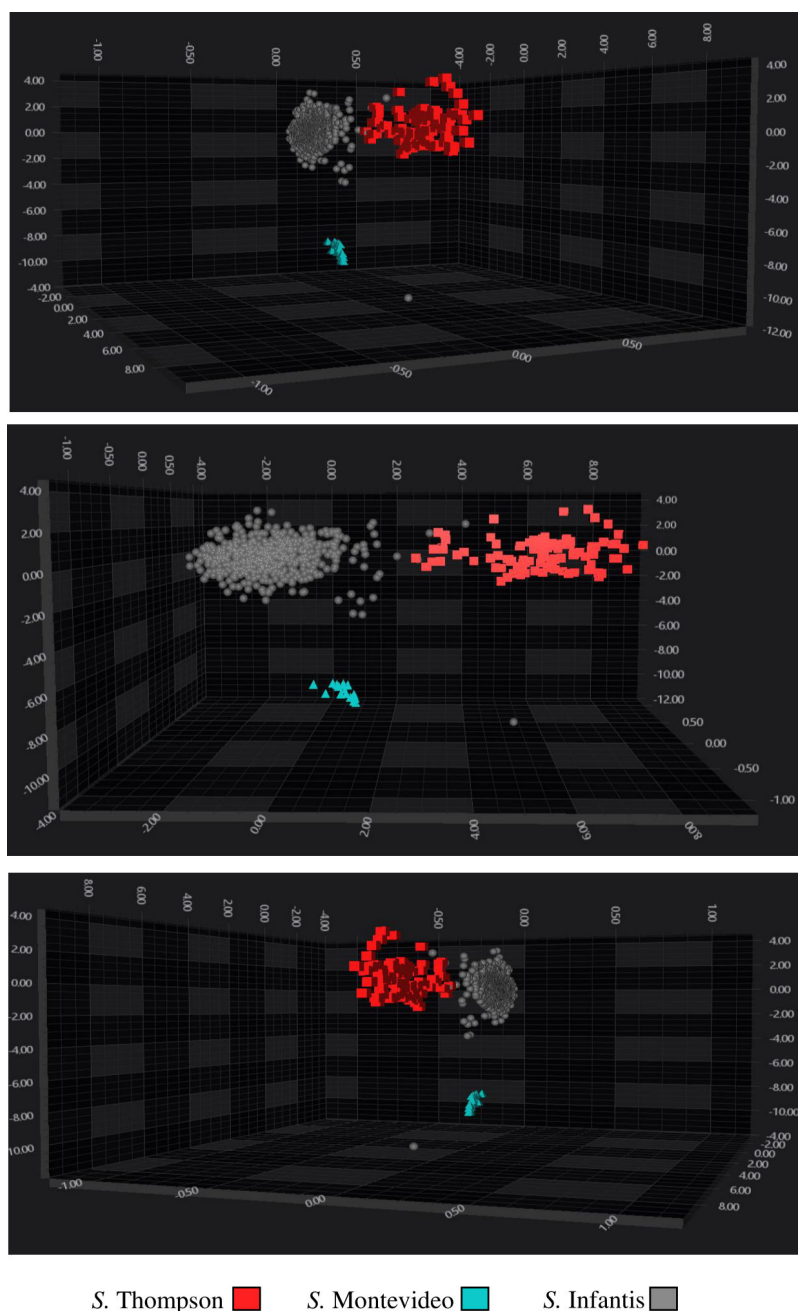


FIG 2 3D scatter plot of the third training performed, showing the distribution of serovars of *Salmonella* serogroup O:7 species in IR spectral space. A total of 30 PCs were used to create a LDA by serovar. The spectra are colored according to the serovars: red squares: *S. Thompson*; cyan triangles: *S. Montevideo*; gray circles: *S. Infantis*. Every symbol represents a different isolate.

S. Agona has the best-defined cluster, the validation confusion matrix indicates that *S. Abony* had the best overall performance, as it correctly identified all its spectra, despite being closely grouped with *S. Typhimurium*/I,4,[5],12:i:- in the LDA analysis (Fig. 1; Tables 2 and 3). The *S. Agona* classifier, though well-defined in LDA, failed to identify approximately 1% of the total spectra. Meanwhile, the *S. Typhimurium*/I,4,[5],12:i:- classifier had the lowest performance, misclassifying or failing to identify around 8% of its spectra (Fig. 1; Table 2).

The serogroup O:7 classifier validation indicated that the *S. Thompson* and *S. Infantis* classifier was the one that exhibited the best performance, correctly identifying a large

TABLE 6 Confusion matrix for the validation of O:9 classifier^a

		Class predicted			
		Panama	Enteritidis	Uncertain	Class recall
Actual class	Panama	20	5	0	80% (20/25)
	Enteritidis	0	289	1	100% (289/290)
	Class precision	100%	98%		

^aAccuracy: 98% (309/315).

number of spectra; this is exhibited in the LDA representation, where the spectra were grouped in a well-identified cluster. Finally, the *S. Montevideo* classifier exhibited the worst performance, misclassifying one-third of the total spectrum, even when the LDA indicated a well-characterized and separated cluster group (Tables 4 and 5; Fig. 2). The small number of *S. Montevideo* isolates and spectra could contribute to the misclassification because the classifier improves predictive performance as the number of training spectra increases isolates.

Analyzing the results obtained for the serogroup O:9 classifier validation, the best performance was obtained by the *S. Enteritidis* classifier, correctly identifying almost 100% of spectra. The performance exhibited by the *S. Panama* classifier is acceptable based on the number of total spectra utilized for the training set, compared with the *S. Enteritidis* classifier (Tables 6 and 7).

As we can evidence in our results, the best performance was not always reached by the classifier that was developed utilizing a higher number of spectra, like some classifiers that exhibited high performance but a lowest number of spectra in the training set; on the other hand, we also evidence that some classifiers, developed with the high number of spectra, also exhibited high performance. This result suggests that the number of required spectra to develop a high-accuracy classifier may differ by each strain type. Additionally, the optimal size for training an SVM model remains uncertain, with no consensus in the existing literature regarding the recommended data size. Recommendations range from hundreds (40) to thousands (28, 29). Based on these suggestions, the number of spectra considered in our study falls within the acceptable range for training and validation data set sizes. Despite this, the investigation could benefit from additional data, as this could further improve the discriminatory power more accurately and robustly by including data processing techniques, such as removal of outlier data and spectra reprocessing (41, 42). We maintain that proper data processing practices, coupled with continuous training and the development of new hierarchical typing algorithms to improve discriminatory power by minimizing the number of trainable features, may have a positive impact on the performance of newly developed classifiers.

Conclusions

Based on our study, we have learned valuable insights into the application of FT-IR technology for *Salmonella* serotyping. Our findings demonstrate its effectiveness in discriminating between *Salmonella* serogroups O:9, O:7, and O:4, including key serovars such as Typhimurium, Enteritidis, and Infantis. To maximize serovar typing accuracy, we recommend focusing the training set on serovars within the same serogroup and minimizing isolate diversity. In our study, successful serovar identification was achieved, underscoring the importance of adequate training data. We encountered limitations during our research about the number of isolates available by each desired serovar: some of them were represented by under 10 isolates and others above a hundred isolates.

TABLE 7 Values of sensitivity, specificity, PPV, and NPV for the validation per serogroup O:9

	Sensitivity	Specificity	PPV	NPV
<i>S. Panama</i>	80%	100%	100%	98%
<i>S. Enteritidis</i>	100%	80%	98%	100%

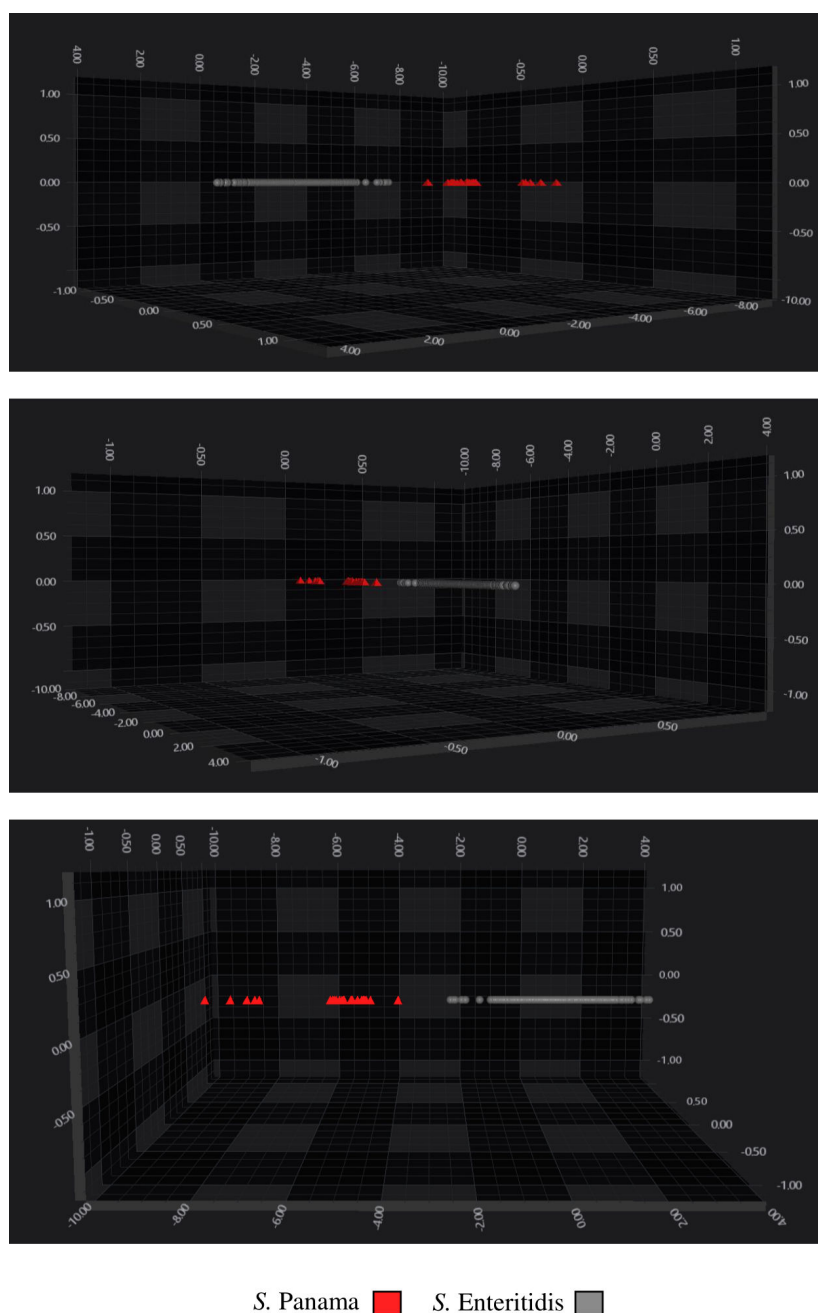


FIG 3 3D scatter plot of the third training performed, showing the distribution of serovars of *Salmonella* serogroup O:9 species in IR spectral space. A total of 30 PCs were used to create a LDA by serovar. The spectra are colored according to the serovars: red triangles: *S. Panama*; gray circles: *S. Enteritidis*. Every symbol represents a different isolate.

Interestingly, the number of isolates by serovar is not directly related to the accuracy of the developed classifier.

Accurate identification of *Salmonella* serovars has far-reaching implications across multiple sectors, including clinical diagnostics, epidemiological surveillance, and various branches of the food industry. By improving the speed and accuracy of serovar identification, advanced technologies like the IR Biotyper provide solutions to long-standing challenges in public health and industrial settings.

To maximize the impact of these technologies, we recommend ongoing validation, training data set updates, and adaptation to local and emerging *Salmonella* serovars. By

doing so, advanced serotyping methods can remain reliable, scalable, and effective in addressing public health challenges, supporting surveillance efforts and improving food safety practices across industries.

These applications underscore the versatility and potential of advanced serotyping technologies to bridge the gap between scientific innovation and practical solutions in diagnostics, surveillance, and industry.

ACKNOWLEDGMENTS

This study was partially supported by the Agencia de Investigación y Desarrollo de Chile (ANID) through FONDEF IDeA I+D 2024 Folio 24110291, FONDECYT de Postdoctorado Folio 3230796, FONDECYT Regular grant number 1231082, FONDECYT Regular grant number 1221536, and FIC Regional Government O'Higgins Region through the Competitiveness Innovation Fund Program (FIC) 2020 grant number IDI 40027681-0.

We thank to Servicio Agrícola Ganadero (SAG) for providing *S. Agona* isolates and Dr. Angélica Reyes Jara for providing *S. Typhimurium* and 4,[5],12:i:- isolates from her collection at Universidad de Chile.

D.F-G.: conceptualization, data curation, formal analysis, validation, and writing—original draft; J.J-R.: conceptualization, data curation, formal analysis, methodology, and validation; A.P-I.: conceptualization, validation, and writing—original draft; P.C-D.: conceptualization and validation; T.G-V.: methodology; F.D.: methodology, A.D.A.: writing—review; A.M-S.: writing—review and editing; and P.G.:conceptualization, funding acquisition, project administration, resources, supervision, writing—review and editing. All authors have read and agreed to the published version of the manuscript.

During the preparation of this work, the authors used Grammarly and ChatGPT to improve readability and language. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

AUTHOR AFFILIATIONS

¹Escuela de Medicina Veterinaria, Facultad de Medicina, Pontificia Universidad Católica de Chile, Santiago, Chile

²Departamento de Laboratorios Clínicos, Escuela de Medicina, Facultad de Medicina, Pontificia Universidad Católica de Chile, Santiago, Chile

³Red de Salud UC-CHRISTUS, Laboratorio de Microbiología, Santiago, Chile

⁴Escuela de Medicina Veterinaria, Centro de Estudios e Investigación en Salud y Sociedad, Facultad de Ciencias Médicas, Universidad Bernardo O'Higgins, Santiago, Chile

⁵Escuela de Medicina Veterinaria, Facultad de Ciencias de la vida, Universidad Andrés Bello, Santiago, Chile

AUTHOR ORCIDs

Diego Fredes-García  <http://orcid.org/0009-0002-5031-9821>

Javiera Jiménez-Rodríguez  <http://orcid.org/0009-0007-9203-5583>

Alejandro Piña-Iturbe  <http://orcid.org/0000-0001-6190-9494>

Fernando Dueñas  <http://orcid.org/0000-0001-6424-1960>

Aniela Wozniak  <http://orcid.org/0000-0001-9559-7660>

Andrea I. Moreno-Switt  <http://orcid.org/0000-0001-8231-9397>

Patricia García  <http://orcid.org/0000-0002-5983-2392>

AUTHOR CONTRIBUTIONS

Diego Fredes-García, Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Visualization, Writing – original draft | Javiera Jiménez-Rodríguez, Conceptualization, Data curation, Formal analysis, Methodology, Software, Validation, Writing – original draft | Alejandro Piña-Iturbe, Conceptualization, Data curation,

Investigation, Methodology, Supervision, Writing – original draft | Pablo Caballero-Díaz, Conceptualization, Writing – original draft | Tamara González-Villarroel, Data curation, Methodology | Fernando Dueñas, Data curation, Methodology | Aniela Wozniak, Conceptualization, Investigation, Supervision, Writing – review and editing | Aiko D. Adell, Conceptualization, Investigation, Supervision, Writing – review and editing | Andrea I. Moreno-Switt, Conceptualization, Investigation, Supervision, Writing – review and editing | Patricia García, Conceptualization, Funding acquisition, Project administration, Resources, Supervision, Writing – review and editing

ETHICS APPROVAL

Ethical clearance was approved under the ID 240621037.

ADDITIONAL FILES

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

Supplemental Material

Table S1 (Spectrum00159-25-s0001.xlsx). *Salmonella* O:4, O:7, and O:9 isolates by different sources with serovar and serogroup classification.

REFERENCES

- Ikuta KS, Swetschinski LR, Robles Aguilar G, Sharara F, Mestrovic T, Gray AP, Davis Weaver N, Wool EE, Han C, Gershberg Hayoon A, et al. 2022. Global mortality associated with 33 bacterial pathogens in 2019: a systematic analysis for the Global Burden of Disease Study 2019. *The Lancet* 400:2221–2248. [https://doi.org/10.1016/S0140-6736\(22\)02185-7](https://doi.org/10.1016/S0140-6736(22)02185-7)
- World Health Organization (WHO). 2024. Bacterial priority pathogens list. Bacterial pathogens of public health importance to guide research, development and strategies to prevent and control antimicrobial resistance. World Health Organization
- Chattaway MA, Langridge GC, Wain J. 2021. *Salmonella* nomenclature in the genomic era: a time for change. *Sci Rep* 11:7494. <https://doi.org/10.1038/s41598-021-86243-w>
- Grimont PAD, Weill FX, WHO collaborating centre for reference and research on *Salmonella*. 2007. Antigenic formulae of the *Salmonella* serovars. 9th ed, p 1–166. World Health Organisation.
- 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>
- European Food Safety Authority (EFSA), European Centre for Disease Prevention and Control (ECDC). 2023. The European union one health 2022 zoonoses report. *EFSA J* 21:e8442. <https://doi.org/10.2903/j.efsa.2023.8442>
- Katz TS, Harhay DM, Schmidt JW, Wheeler TL. 2024. Identifying a list of *Salmonella* serotypes of concern to target for reducing risk of salmonellosis. *Front Microbiol* 15:1307563. <https://doi.org/10.3389/fmicb.2024.1307563>
- Gong J, Zeng X, Zhang P, Zhang D, Wang C, Lin J. 2019. Characterization of the emerging multidrug-resistant *Salmonella enterica* serovar Indiana strains in China. *Emerg Microbes Infect* 8:29–39. <https://doi.org/10.1080/22221751.2018.1558961>
- Liao Y-S, Chen B-H, Hong Y-P, Teng R-H, Wang Y-W, Liang S-Y, Liu Y-Y, Tu Y-H, Chen Y-S, Chang J-H, Tsao C-S, Chiou C-S. 2019. Emergence of multidrug-resistant *Salmonella enterica* serovar goldcoast strains in Taiwan and international spread of the ST358 clone. *Antimicrob Agents Chemother* 63:e01122-19. <https://doi.org/10.1128/AAC.01122-19>
- Ford L, Ellison Z, Schwensohn C, Griffin I, Birhane MG, Cote A, Fortenberry GZ, Tecle S, Higa J, Spencer S, Patton B, Patel J, Dow J, Maroufi A, Robbins A, Donovan D, Fitzgerald C, Burrell S, Tolar B, Folster JP, Cooley LA, Francois Watkins LK. 2023. Strain of multidrug-resistant *Salmonella* Newport remains linked to travel to Mexico and U.S. beef products-United States. *MMWR Morb Mortal Wkly Rep* 72:1225–1229. <https://doi.org/10.15585/mmwr.mm7245a3>
- Aworh MK, Nilsson P, Egyir B, Owusu FA, Hendriksen RS. 2024. Rare serovars of non-typhoidal *Salmonella enterica* isolated from humans, beef cattle and abattoir environments in Nigeria. *PLoS One* 19:e0296971. <https://doi.org/10.1371/journal.pone.0296971>
- Piña-Iturbe A, Díaz-Gavidia C, Álvarez FP, Barron-Montenegro R, Álvarez-Espejo DM, García P, Solís D, Constenla-Albornoz R, Toro M, Olivares-Pacheco J, Reyes-Jara A, Meng J, Bell RL, Moreno-Switt AI. 2024. Genomic characterisation of the population structure and antibiotic resistance of *Salmonella enterica* serovar Infantis in Chile, 2009-2022. *Lancet Reg Health Am* 32:100711. <https://doi.org/10.1016/j.lana.2024.100711>
- Silva C, Calva E, Maloy S. 2014. One health and food-borne disease: *Salmonella* transmission between humans, animals, and plants. *Microbiol Spectr* 2:OH-0020 <https://doi.org/10.1128/microbiolspec.OH-0020-2013>
- Wattiau P, Boland C, Bertrand S. 2011. Methodologies for *Salmonella enterica* subsp. *enterica* subtyping: gold standards and alternatives. *Appl Environ Microbiol* 77:7877–7885. <https://doi.org/10.1128/AEM.05527-11>
- Cordovana M, Mauder N, Join-Lambert O, Gravey F, LeHello S, Auzou M, Pitti M, Zoppi S, Buhl M, Steinmann J, et al. 2022. Machine learning-based typing of *Salmonella enterica* O-serogroups by the Fourier-transform infrared (FTIR) spectroscopy-based IR biotyper system. *J Microbiol Methods* 201:106564. <https://doi.org/10.1016/j.mimet.2022.106564>
- Diep B, Barretto C, Portmann A-C, Fournier C, Karczmarek A, Voets G, Li S, Deng X, Klijn A. 2019. *Salmonella* serotyping; comparison of the traditional method to a microarray-based method and an *in silico* platform using whole genome sequencing data. *Front Microbiol* 10:2554. <https://doi.org/10.3389/fmicb.2019.02554>
- Yoshida C, Gurnik S, Ahmad A, Blimkie T, Murphy SA, Kropinski AM, Nash JHE. 2016. Evaluation of molecular methods for identification of *Salmonella* serovars. *J Clin Microbiol* 54:1992–1998. <https://doi.org/10.1128/JCM.00262-16>
- Chui L, Ferrato C, Li V, Christianson S. 2021. Comparison of molecular and *in silico* *Salmonella* serotyping for *Salmonella* surveillance. *Microorganisms* 9:955. <https://doi.org/10.3390/microorganisms9050955>
- Jauneikaite E, Baker KS, Nunn JG, Midega JT, Hsu LY, Singh SR, Halpin AL, Hopkins KL, Price JR, Srikanthiah P, Egyir B, Okeke IN, Holt KE, Peacock SJ, Feasey NA, SEDRIC Genomics Surveillance Working Group. 2023. Genomics for antimicrobial resistance surveillance to support infection prevention and control in health-care facilities. *Lancet Microbe* 4:e1040–e1046. [https://doi.org/10.1016/S2666-5247\(23\)00282-3](https://doi.org/10.1016/S2666-5247(23)00282-3)
- Lilo T, Morais CLM, Shenton C, Ray A, Gurusinghe N. 2022. Revising Fourier-transform infrared (FT-418 IR) and Raman spectroscopy towards

- brain cancer detection. Photodiagnosis Photodyn Ther 38:102785. <https://doi.org/10.1016/j.pdpdt.2022.102785>
21. Baker MJ, Trevisan J, Bassan P, Bhargava R, Butler HJ, Dorling KM, Fielden PR, Fogarty SW, Fullwood NJ, Heys KA, Hughes C, Lasch P, Martin-Hirsch PL, Obinaju B, Sockalingum GD, Sulé-Suso J, Strong RJ, Walsh MJ, Wood BR, Gardner P, Martin FL. 2014. Using fourier transform IR spectroscopy to analyze biological materials. Nat Protoc 9:1771–1791. <https://doi.org/10.1038/nprot.2014.110>
 22. Muchaamba F, Stephan R. 2024. A comprehensive methodology for microbial strain typing using fourier-transform infrared spectroscopy. Methods Protoc 7:48. <https://doi.org/10.3390/mps7030048>
 23. Naumann D, Helm D, Labischinski H. 1991. Microbiological characterizations by FT-IR spectroscopy. Nature 351:81–82. <https://doi.org/10.1038/351081a0>
 24. Park S, Ryoo N. 2024. Comparative analysis of IR-biotyper, MLST, cgMLST, and WGS for clustering of vancomycin-resistant *Enterococcus faecium* in a neonatal intensive care unit. Microbiol Spectr 12:e0411923. <https://doi.org/10.1128/spectrum.04119-23>
 25. Nakajima N, Jinnai M, Izumiyama S, Kuroki T. 2024. Evaluating fourier-transform infrared spectroscopy with IR biotyper as a faster and simpler method for investigating the sources of an outbreak of legionellosis. Eur J Clin Microbiol Infect Dis 43:991–997. <https://doi.org/10.1007/s10096-024-04781-3>
 26. Hu Y, Zhou H, Lu J, Sun Q, Liu C, Zeng Y, Zhang R. 2021. Evaluation of the IR biotyper for *Klebsiella pneumoniae* typing and its potentials in hospital hygiene management. Microb Biotechnol 14:1343–1352. <https://doi.org/10.1111/1751-7915.13709>
 27. Rakovitsky N, Frenk S, Kon H, Schwartz D, Temkin E, Solter E, Paikin S, Cohen R, Schwaber MJ, Carmeli Y, Lellouche J. 2020. Fourier transform infrared spectroscopy is a new option for outbreak investigation: a retrospective analysis of an extended-spectrum-beta-lactamase-producing *Klebsiella pneumoniae* outbreak in a neonatal intensive care unit. J Clin Microbiol 58:e00098-20. <https://doi.org/10.1128/JCM.00098-20>
 28. Passaris I, Mauder N, Kostrzewa M, Burckhardt I, Zimmermann S, van Sorge NM, Slotved H-C, Desmet S, Ceyssens P-J. 2022. Validation of Fourier transform infrared spectroscopy for serotyping of *Streptococcus pneumoniae*. J Clin Microbiol 60:e0032522. <https://doi.org/10.1128/jcm.00325-22>
 29. Uribe G, Salipante SJ, Curtis L, Lieberman JA, Kurosawa K, Cookson BT, Hoogestraat D, Stewart MK, Olmstead T, Bourassa L. 2023. Evaluation of Fourier transform-infrared spectroscopy (FT-IR) as a control measure for nosocomial outbreak investigations. J Clin Microbiol 61:e0034723. <https://doi.org/10.1128/jcm.00347-23>
 30. BRUKER. 2023. Microbiology & diagnostics microbial strain typing. Available from: <https://www.bruker.com/en/products-and-solutions/microbiology-and-diagnostics/microbial-strain-typing.html>
 31. Campos J, Sousa C, Mourão J, Lopes J, Antunes P, Peixe L. 2018. Discrimination of non-typhoid *Salmonella* serogroups and serotypes by Fourier transform infrared spectroscopy: a comprehensive analysis. Int J Food Microbiol 285:34–41. <https://doi.org/10.1016/j.ijfoodmicro.2018.07.005>
 32. Cordovana M, Mauder N, Kostrzewa M, Wille A, Rojak S, Hagen RM, Ambretti S, Pongolini S, Soliani L, Justesen US, Holt HM, Join-Lambert O, Hello SL, Auzou M, Veloo AC, May J, Frickmann H, Dekker D. 2021. Classification of *Salmonella enterica* of the (Para)-typhoid fever group by Fourier-transform infrared (FTIR) spectroscopy. Microorganisms 9:853. <https://doi.org/10.3390/microorganisms9040853>
 33. Toro M, Weller D, Ramos R, Díaz L, Alvarez FP, Reyes-Jara A, Moreno-Switt AI, Meng J, Adell AD. 2022. Environmental and anthropogenic factors associated with the likelihood of detecting *Salmonella* in agricultural watersheds. Environ Pollut 306:119298. <https://doi.org/10.1016/j.envpol.2022.119298>
 34. Rivera D, Allel K, Dueñas F, Tardone R, Soza P, Hamilton-West C, Moreno-Switt AI. 2021. Screening the presence of non-typhoidal *Salmonella* in different animal systems and the assessment of antimicrobial resistance. Animals (Basel) 11:1532. <https://doi.org/10.3390/ani11061532>
 35. Soza-Ossandón P, Rivera D, Tardone R, Riquelme-Neira R, García P, Hamilton-West C, Adell AD, González-Rocha G, Moreno-Switt AI. 2020. Widespread environmental presence of multidrug-resistant *Salmonella* in an equine veterinary hospital that received local and international horses. Front Vet Sci 7:346. <https://doi.org/10.3389/fvets.2020.00346>
 36. Alegria-Moran R, Lazo A, Rivera D, Toledo V, Moreno-Switt A, Hamilton-West C. 2017. First detection of *Salmonella* spp. in backyard production farms from central Chile. Online J Public Health Inform 9. <https://doi.org/10.5210/ojphi.v9i1.7744>
 37. Piña-Iturbe A, Fredes-García D, García P, Porte L, Johnson TJ, Singer RS, Toro M, Munita JM, Moreno-Switt AI. 2025. *Salmonella enterica* serovar abony outbreak caused by clone of reference strain WDCM 00029, Chile, 2024. Emerg Infect Dis 31:186–189. <https://doi.org/10.3201/eid3101.241012>
 38. Ranieri ML, Shi C, Moreno Switt AI, den Bakker HC, Wiedmann M. 2013. Comparison of typing methods with a new procedure based on sequence characterization for *Salmonella* serovar prediction. J Clin Microbiol 51:1786–1797. <https://doi.org/10.1128/JCM.03201-12>
 39. Männig A, Baldauf NA, Rodríguez-Romo LA, Yousef AE, Rodríguez-Saona LE. 2008. Differentiation of *Salmonella enterica* serovars and strains in cultures and food using infrared spectroscopic and microspectroscopic techniques combined with soft independent modeling of class analogy pattern recognition analysis. J Food Prot 71:2249–2256. <https://doi.org/10.4315/0362-028X-71.11.2249>
 40. Barker KR, Santino M, LiPuma JJ, Tullis E, Muller MP, Matukas LM, Tadors M. 2021. Fourier transform infrared spectroscopy for typing *Burkholderia cenocepacia* ET12 isolates. Microbiol Spectr 9:e0183121. <https://doi.org/10.1128/Spectrum.01831-21>
 41. Chicco D. 2017. Ten quick tips for machine learning in computational biology. BioData Min 10:35. <https://doi.org/10.1186/s13040-017-0155-3>
 42. Pascale MR, Bisognin F, Mazzotta M, Girolamini L, Marino F, Dal Monte P, Cordovana M, Scaturro M, Ricci ML, Cristino S. 2022. Use of Fourier-transform infrared spectroscopy with IR biotyper system for *Legionella pneumophila* serogroups identification. Front Microbiol 13:866426. <https://doi.org/10.3389/fmicb.2022.866426>