Machine learning-based typing of Salmonella enterica O-serogroups by the Fourier-Transform Infrared (FTIR) Spectroscopy-based IR Biotyper system

Cordovana, Miriam; Mauder, Norman; Join-Lambert, Olivier; Gravey, François; LeHello, Simon; Auzou, Michel; Pitti, Monica; Zoppi, Simona; Buhl, Michael; Steinmann, Joerg

Published in:
Journal of microbiological methods

DOI:
10.1016/j.mimet.2022.106564

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2022

Journal of Microbiological Methods 201 (2022) 106564

Contents lists available at ScienceDirect

Journal of Microbiological Methods

journal homepage: www.elsevier.com/locate/jmicmeth

Machine learning-based typing of *Salmonella enterica* O-serogroups by the Fourier-Transform Infrared (FTIR) Spectroscopy-based IR Biotyper system

Miriam Cordovana [a,*], Norman Mauder [a], Olivier Join-Lambert [b], François Gravey [b], Simon LeHello [b], Michel Auzou [b], Monica Pitti [c], Simona Zoppi [c], Michael Buhl [d], Joerg Steinmann [d], Hagen Frickmann [e,f], Denise Dekker [g], Yumiko Funashima [h], Zenzo Nagasawa [h], József Soki [i], László Orosz [i], Alida C. Veloo [j], Ulrik S. Justesen [k], Hanne M. Holt [k], Andrea Liberatore [l], Simone Ambretti [l], Stefano Pongolini [m], Laura Soliani [m], Andreas Wille [n], Sandra Rojak [o], Ralf Matthias Hagen [o], Jürgen May [g,p], A.B. Pranada [q], Markus Kostrzewa [a]

[a] *Bruker Daltonics GmbH & Co. KG, 28359 Bremen, Germany*
[b] *Université de Caen, Normandie, 14032, Cedex 5, Caen, France*
[c] *Istituto Zooprofilattico Sperimentale del Piemonte Liguria e Valle d'Aosta, (SS Patologia Animale – Simona; SS Microbiologia Comparativa Specialistica - Centro di Riferimento Tipizzazione delle Salmonelle (CeRTiS) – Monica), via Bologna 148,Torino, Italy*
[d] *Institute of Clinical Hygiene, Medical Microbiology and Infectiology, Paracelsus Medical University, Nuremberg, Germany*
[e] *Department of Microbiology and Hospital Hygiene, Bundeswehr Hospital Hamburg, 20359 Hamburg, Germany*
[f] *Institute for Medical Microbiology, Virology and Hygiene, University Medicine Rostock, 18057 Rostock, Germany*
[g] *Bernhard Nocht Institute for Tropical Medicine Hamburg, 20359 Hamburg, Germany*
[h] *Department of Medical Technology and Sciences, School of Health Sciences, Fukuoka International University of Health and Welfare, 137-1 Enokizu, Okawa, Fukuoka, Japan*
[i] *Institute of Medical Microbiology, Albert Szent-Györgyi Health Centre and Medical School, University of Szeged, Szeged, Hungary*
[j] *Department of Medical Microbiology and Infection Prevention, University Medical Center Groningen, University of Groningen, 9700 AB Groningen, the Netherlands*
[k] *Department of Clinical Microbiology, Odense University Hospital, 5000 Odense C, Denmark*
[l] *Operative Unit of Microbiology, IRCCS-Azienda Ospedaliero Policlinico Sant'Orsola-Universitaria di Bologna, 40138 Bologna, Italy*
[m] *Risk Analysis and Genomic Epidemiology Unit, Istituto Zooprofilattico Sperimentale della Lombardia e dell'Emilia-Romagna, 43126, Italy*
[n] *Institute for Hygiene and Environment, City of Hamburg, 20539 Hamburg, Germany*
[o] *Department of Microbiology and Hospital Hygiene, Bundeswehr Central Hospital Koblenz, 56070 Koblenz, Germany*
[p] *University Medical Center Hamburg-Eppendorf (UKE), Tropical Medicine II, Hamburg, Germany*
[q] *Department of Medical Microbiology, MVZ Dr. Eberhard & Partner Dortmund, Dortmund, Balkenstrasse 17-19, 44137 Dortmund, Germany*

ARTICLE INFO

Keywords:
Salmonella enterica
FT-IR spectroscopy
FTIRS
Salmonella typing
IR Biotyper
Machine learning

ABSTRACT

Background: Salmonella enterica is among the major burdens for public health at global level. Typing of salmonellae below the species level is fundamental for different purposes, but traditional methods are expensive, technically demanding, and time-consuming, and therefore limited to reference centers. Fourier transform infrared (FTIR) spectroscopy is an alternative method for bacterial typing, successfully applied for classification at different infra-species levels.

Aim: This study aimed to address the challenge of subtyping *Salmonella enterica* at O-serogroup level by using FTIR spectroscopy. We applied machine learning to develop a novel approach for *S. enterica* typing, using the

* Corresponding author.
E-mail addresses: miriam.cordovana@bruker.com (M. Cordovana), norman.mauder@bruker.com (N. Mauder), olivier.join-lambert@unicaen.fr (O. Join-Lambert), Lehello-s@chu-caen.fr (S. LeHello), auzou-m@chu-caen.fr (M. Auzou), monica.pitti@izsto.it (M. Pitti), simona.zoppi@izsto.it (S. Zoppi), Michael.Buhl@klinikum-nuernberg.de (M. Buhl), Joerg.Steinmann@klinikum-nuernberg.de (J. Steinmann), frickmann@bnitm.de, hagen.frickmann@med.uni-rostock.de (H. Frickmann), dekker@bnitm.de (D. Dekker), funashima@iuhw.ac.jp (Y. Funashima), nagasa@iuhw.ac.jp (Z. Nagasawa), soki.jozsef@med.u-szeged.hu (J. Soki), orosz.laszlo@med.u-szeged.hu (L. Orosz), a.c.m.veloo@umcg.nl (A.C. Veloo), Ulrik.Stenz.Justesen@rsyd.dk (U.S. Justesen), Hanne.Holt@rsyd.dk (H.M. Holt), andrea.liberatore@studio.unibo.it (A. Liberatore), simone.ambretti@aosp.bo.it (S. Ambretti), stefano.pongolini@izsler.it (S. Pongolini), laura.soliani@izsler.it (L. Soliani), andreas.wille@hu.hamburg.de (A. Wille), sandrarojak@bundeswehr.org (S. Rojak), ralfmatthiashagen@bundeswehr.org (R.M. Hagen), may@bnitm.de, j.may@uke.de (J. May), apranada@labmed.de (A.B. Pranada), markus.kostrzewa@bruker.com (M. Kostrzewa).

https://doi.org/10.1016/j.mimet.2022.106564
Received 14 June 2022; Received in revised form 30 August 2022; Accepted 30 August 2022
Available online 6 September 2022

FTIR-based IR Biotyper® system (IRBT; Bruker Daltonics GmbH & Co. KG, Germany). We investigated a multicentric collection of isolates, and we compared the novel approach with classical serotyping-based and molecular methods.

Methods: A total of 958 well characterized *Salmonella* isolates (25 serogroups, 138 serovars), collected in 11 different centers (in Europe and Japan), from clinical, environmental and food samples were included in this study and analyzed by IRBT. Infrared absorption spectra were acquired from water-ethanol bacterial suspensions, from culture isolates grown on seven different agar media. In the first part of the study, the discriminatory potential of the IRBT system was evaluated by comparison with reference typing method/s. In the second part of the study, the artificial intelligence capabilities of the IRBT software were applied to develop a classifier for *Salmonella* isolates at serogroup level. Different machine learning algorithms were investigated (artificial neural networks and support vector machine). A subset of 88 pre-characterized isolates (corresponding to 25 serogroups and 53 serovars) were included in the training set. The remaining 870 samples were used as validation set. The classifiers were evaluated in terms of accuracy, error rate and failed classification rate.

Results: The classifier that provided the highest accuracy in the cross-validation was selected to be tested with four external testing sets. Considering all the testing sites, accuracy ranged from 97.0% to 99.2% for non-selective media, and from 94.7% to 96.4% for selective media.

Conclusions: The IRBT system proved to be a very promising, user-friendly, and cost-effective tool for *Salmonella* typing at serogroup level. The application of machine learning algorithms proved to enable a novel approach for typing, which relies on automated analysis and result interpretation, and it is therefore free of potential human biases. The system demonstrated a high robustness and adaptability to routine workflows, without the need of highly trained personnel, and proving to be suitable to be applied with isolates grown on different agar media, both selective and unselective. Further tests with currently circulating clinical, food and environmental isolates would be necessary before implementing it as a potentially stand-alone standard method for routine use.

1. Introduction

Salmonella enterica is one of the leading causes of foodborne diseases worldwide, representing a major public health burden for both low-income and industrialized countries. It is responsible of 180 million cases of salmonellosis, up to 24.2 million cases of typhoid fever, and 298,000 estimated deaths per year (Antillón et al., 2017; CDC, 2017; EFSA and ECDC, 2017; WHO, 2016). *S. enterica* is transmitted to humans by consumption of a wide range of contaminated foods, thereby it is involved both in endemic and epidemic scenarios (Antunes et al., 2016, 2017; EFSA and ECDC, 2017; Mourão et al., 2014; Painter et al., 2013). Differentiation of *S. enterica* at subspecies level is crucial for epidemiological investigations and for the control of foodborne outbreaks, as well as for the clinical management of infections. Many methods are used for *S. enterica* typing (serotyping, phage typing, DNA-based methods) (Tang et al., 2019), which allow subspecies discrimination at different levels, but these are laborious and cost-intensive, and often require high technical expertise (Sabat et al., 2013). Globally, serotyping, based on the agglutination reaction with specific antisera targeting the somatic O-antigen and flagellar H-antigens (Grimont and Weill, 2007), is still the recognized and most widely used approach to classify *S. enterica*. Discrimination at serogroup level is the most frequently applied phenotypic method, since few serotypes, belonging to serogroups D, B, C and E, namely *S.* Eenteritidis, *S.* Ttyphimurium and its monophasic variant (*S.* 1,4,[5],12:i:-), and the emerging *S.* Stanley, *S.* Infantis, *S.* Rissen, *S.* Newport and *S.* Kentucky cause the majority of infections worldwide (CDC, 2017; EFSA and ECDC, 2017; Grimont and Weill, 2007). However, any deeper discrimination at serotype level, despite its potential clinical or epidemiological relevance, is restricted to reference laboratories, since it is complex, expensive, and time-consuming (Parmley et al., 2013).

Fourier Transform Infrared (FTIR) Spectroscopy, traditionally used in analytical chemistry for decades, has also been successfully applied for the discrimination of bacteria at different taxonomic levels (genera, species, serogroup/type, and even at strain level), based on the analysis of the intact microbial cells or of the outer membrane components (Baker et al., 2014; Griffiths and De Haseh, 2007; Helm et al., 1991; Lasch and Naumann, 2015; Naumann et al., 1991). It proved to be a simple, quick, high-throughput and cost-effective technique (Davis and Mauer, 2010; Preisner et al., 2007; Quintelas et al., 2018; Stuart, 2004; Wenning and Scherer, 2013; Zamowiec et al., 2015). *Salmonella enterica* has been shown an interesting and promising bacterial species to be investigated with FTIR spectroscopy, because of its high antigenic diversity and the associated varying clinical relevance. The different length of somatic antigens, and the high carbohydrate diversity of O-units, supposed to have a great impact on the cell surface structure, have the potential to enable a differentiation based on FTIR methodology. Several research groups investigated this approach to discriminate *S. enterica* serotypes using multivariate analysis and different bacterial collections (De Lamo-Castellví et al., 2010; Kim et al., 2006; Männing et al., 2008; Preisner et al., 2010; Sundaram et al., 2012). More recently, the potential of FTIR spectroscopy was assessed by Campos et al. in a study including comprehensive and robust *Salmonella* collections (Campos et al., 2018). The study proved that this methodology represents a reliable and alternative technique for an accurate discrimination of *Salmonella* isolates belonging to B, C, D and E serogroups, C1, C2 and C3, and E1-E2-E3 and E4 subgroups, as well as for a classification of particularly relevant serotypes (*S.* Rissen, *S.* Enteritidis and *S.* Senftenberg). However, further studies are required to provide more thorough molecular assessment-based insights into the potential and limitations of this methodology.

In the present study, the IR Biotyper® system (IRBT, Bruker Daltonics GmbH & Co. KG), an FTIR-based commercially available system for microbial typing, was evaluated for *S. enterica* typing at O-serogroup level. An innovative typing approach was developed, applying artificial intelligence and machine learning (ML). ML uses specific software algorithms to automate computers to make predictions based on biological data. The algorithms learn to identify and recognize the features of the training set. Based on the data they have learnt, they allow the automated classification of unknown samples by application of the marker model calculated on the set of training spectra. To date, several ML techniques are available, well described and established, and they have been applied in almost all disciplines of biological sciences, including medicine. In this study, artificial neural network (ANN) and support vector machine (SVM) algorithms, two of the most widely used ML algorithms, implemented in the IR Biotyper® software, were used to build classifiers for the typing of *Salmonella* species at O-serogroup level. A large collection of strains, isolated from human, food-related, and environmental samples collected at different European sites was included, and a training and the testing sets were defined. The spectra were measured from bacterial cultures on the most widely used culture media. The performance of the classifiers was evaluated in terms of accuracy and error rate, testing different collections of isolates and spectra from different locations. In addition, the impact of the culture

medium on classification accuracy was also evaluated.

2. Materials and methods

2.1. Bacterial collection

Overall, a total of $N = 958$ *Salmonella enterica* non-duplicate isolates of clinical, environmental, veterinary, food-related origins and from culture collections were included in this study. The strains were isolated in/from 11 different hospitals and reference centers, located in different European countries (Université de Caen, Normandie, France; University Medical Center Groningen, the Netherlands; Odense University Hospital, Odense, Denmark; Paracelsus Medical University, Nuremberg, Germany; Institute for Hygiene and Environment, City of Hamburg, Hamburg, Germany; MVZ Dr. Eberhard & Partner Dortmund, Dortmund, Germany; University of Szeged, Szeged, Hungary; University Hospital IRCCS Policlinico Sant'Orsola-Malpighi, Bologna, Italy; Istituto Zooprofilattico Sperimentale della Lombardia e dell'Emilia-Romagna, Parma, Italy; Istituto Zooprofilattico Sperimentale del Piemonte, Liguria e Valle d'Aosta, Torino, Italy; School of Health Sciences, Fukuoka, Japan). In most centers the strains were collected prospectively, without any selection, for a given time frame, while in some cases the isolates were selected from frozen culture collections, with a specific focus (para (typhoidal) serovars, rarer serogroups). Culture collection strains from Leibniz-Institute DSMZ - German Collection of Microorganisms and Cell Cultures (DSMZ) and Collection of the Institute Pasteur (CIP) were also included. Overall, the samples included in the study covered 25 serogroups and 138 serovars. The vast majority ($n = 925$, 96.6%) of isolates belonged to *S. enterica* subsp. *enterica* (subgenus I), while $n = 32$ (3.4%) belonged to other subspecies, namely $n = 12$ subsp. *salamae* (II), $n = 2$ subsp. *arizonae* (IIIa), n = 12 subsp. *diarizonae* (IIIb), $n = 5$ subsp. *houtanae* (IV) and n = 1 subsp. *indica* (VI). Also, one isolate of *Salmonella bongori* (V) was included. The most quantitatively dominating group was O:4 ($n = 422$ strains, 44.1%), followed by O:9 ($n = 186$, 19.4%), O:7 ($n = 111$, 11.6%), O:8 ($n = 60$, 6.3%), O:3,10 ($n = 33$, 3.4%), O:13 ($n = 29$, 3.0%), O:2 (n = 29, 3.0%), O:11 ($n = 21$, 2.2%) and O:28 (n = 12, 1.3%), while the remaining groups were represented by <10 isolates each. (**Table S1**).

All isolates were identified at the genus level in the primary collecting laboratories by MALDI-TOF MS or biochemical methods (API 20E, bioMérieux, Marcy l'Etoile, France). Typing was performed in accordance with the established procedures of each center (serotyping at serotype level (Tang et al., 2019), PFGE (Ribot et al., 2006), PCR for *S.* Typhimurium (Barco et al., 2011; EFSA, 2010; Tennant et al., 2010) or whole genome sequencing). Overall, $n = 881$ isolates were typed at the serovar level, while additional $n = 77$ at the serogroup level.

2.2. Sample preparation

Solid agar cultures for spectra acquisition were incubated at 35 ± 2 °C for 24 ± 1 h in normal atmosphere. Sample preparation for IRBT analysis was performed following manufacturer's instructions. Briefly, a 1 μl overloaded loop with bacterial colony material taken from the confluent part of the plate culture was resuspended in 50 μl of 70% ethanol solution in an IRBT suspension vial. After vortexing, 50 μl of deionized water were added, and the solution mixed by pipetting. Fifteen μl of the bacterial suspension were spotted in three technical replicates onto the 96-spots silicon IRBT target and left to dry for 15–20 min at 35 ± 2 °C in normal atmosphere. The quality controls Infrared Test Standards (IRTS 1 and IRTS 2) of the IRBT kit were resuspended in 90 μl of deionized water, then 90 μl of absolute ethanol were added and mixed. Twelve μl of the suspension were spotted in duplicate onto the IRBT target and left to dry as described for the samples. All steps for IRBT sample preparation and measurements were carried out at a standard laboratory bench, without controlled room temperature and humidity conditions.

The strains tested at Bruker, Bremen, Germany (culture collection strains and isolates collected in the first seven of the above-mentioned centers), were stored in cryovials (Microbanks, PRO-LAB DIAGNOSTICS, Richmond Hill, Canada), and retrieved on Columbia sheep blood agar (CBA - Becton, Dickinson and Company, Sparks, MD, USA), and subcultured on CBA, chocolate agar (CHO), Tryptose Soy agar (TSA), Mueller-Hinton agar (MHA), MacConkey agar, (MacC) (Becton, Dickinson and Company, Sparks, MD, USA), Xylose Lysine Desoxycholate agar and Salmonella-Shigella agar (XLD and SSA – Carl Roth GmbH & Co. KG, Karlsruhe, Germany). Furthermore, a subset of $n = 385$ strains were analyzed from culture on RAPID' Salmonella Medium (RA, Bio-Rad, Marine-la-Coquette, France). Among these, $n = 152$ strains, representing all serogroups and most of serovars, were analyzed also on ChromID® Salmonella Agar (bioMérieux, Marcy-l'Etoile, France).

The strains analyzed at the University Hospital IRCCS Policlinico Sant'Orsola-Malpighi, Bologna, Italy (also including the isolates from Istituto Zooprofilattico Sperimentale della Lombardia e dell'Emilia-Romagna), were retrieved from long-term storage cultures (on TSA, Meus, Piove di Sacco, Italy) and subcultured for the IRBT measurement on Tryptose Soy agar with 5% sheep blood (TSA-SB, Meus). The strains analyzed at the School of Health Sciences of Fukuoka International University of Health and Welfare, Okawa, Japan, were retrieved from frozen cultures onto Salmonella Shigella agar (Eiken Chemical Co., Ltd., Tokyo, Japan), and then subcultured on Mueller-Hinton agar (Eiken Chemical) for IRBT measurement. The strains analyzed at Istituto Zooprofilattico Sperimentale del Piemonte Liguria e Valle d'Aosta, Torino, Italy, were retrieved from cryovials on CBA (Becton Dickinson and directly analyzed by IRBT.

All samples were analyzed in one independent biological replicate, with the exception of CBA in Bremen, from which three independent cultures on three different days were tested.

2.3. Spectra acquisition and analysis

IRBT spectra were acquired in three centers: the Bruker Daltonics bacteriology laboratory in Bremen, Germany ($n = 552$ isolates), Bologna and Torino, Italy ($n = 327$ and $n = 36$ isolates, respectively) and Okawa, Japan ($n = 39$). Spectra acquisition was performed in transmission mode in the spectral range 4000–500 cm^{-1} (mid-IR) using the IRBT spectrometer and OPUS software (Bruker Optics GmbH & Co. KG). Processing and visualization of spectra was performed with the IR Biotyper Client Software (Bruker Daltonics), applying the versions available at the moment of the measurement (V2.1, V.3.0, or V3.1) and using default settings recommended by the manufacturer. After spectra smoothing using the Savitzky-Golay algorithm over 9 data points, the second derivative of the spectra was calculated. Spectra were then cut to 1300–800 cm^{-1} [14] and vector-normalized to correct for preparation-related variance of biomass and hence absorption.

IRTS 1 and IRTS 2 were measured as quality control prior to sample spectra acquisition, in each run. All spectra were acquired intercalating a background spectrum between each sample/control measurement.

2.4. Exploratory unsupervised and supervised multivariate analysis

IRBT data analysis was performed in Bremen. Principal components analysis (PCA) and linear discriminant analysis (LDA) were applied to the whole dataset of isolates measured in Bremen (n = 552), for a first investigation of the clustering capability and the discriminatory power of IRBT for *Salmonella* at O serogroup level. PCA and LDA were also used to estimate the degree of heterogeneity within the most numerous serogroups, especially regarding the most common and most clinically relevant serogroups (O:2, O:4, O:7, O:8, O:9, O:3,10, O:1,3,19, O:13).

2.5. Machine learning and development of automated classifiers

IRBT classifiers consist of a machine learning algorithm (ML),

presently an artificial neural network (ANN) or a support vector machine (SVM), and an outlier detector (OD). The ML is trained with a set of well characterized isolates, to recognize the specific characteristics of each class (O-serogroups). Based on the discriminatory features that it has "learned" during the training, the classifier assigns the unknown samples to one of the predefined classes (O-serogroups). Thus, the classification process represents the attribution of unknown *Salmonella* isolates to one of the O-serogroups included in the training dataset, according to the model calculated by the algorithm. In addition, the OD determines the spectral distance of a sample from the training set and can be used to evaluate similarity of unknown samples with training samples, as well as to detect and disregard unrelated samples. The outlier value enables to deliver a classification result with a "traffic light" color code scoring system, which depicts the reliability of the classification. The cut-off values applied to define the outlier categories are extrapolated from the distribution of the outlier values of the validation cohort of samples (samples not included in the training set), considering the Youden index. A "green score" result means that the sample spectrum is located within the spectral space of the training set, therefore it can be considered highly reliable. A "yellow score" result means that the sample spectrum is located at the periphery of the spectral space of the training set, therefore it can be considered moderately reliable. A "red score" value means that the sample spectrum is located far from the samples included in the training set, therefore it cannot be considered reliable, as the isolate could either not belong to any known class included in the training set (in this case, another O-serogroup, or also an unknown serotype of a known serogroup), or the acquisition/incubation conditions differ too much from the ones in the training set.

In this study, different versions of a *Salmonella* O-group classifier were built applying the above-mentioned ML algorithms to the same training set. Presently, ANN and SVM with either linear or Gaussian radial basis function (RBF) kernel are available algorithms in the IRBT software. During the training of the classifiers, a 4-fold cross validation was performed automatically to assess accuracy (= true positives / all classifications) and check for overfitting. The parameters of the classification algorithms (PCs used, number of training cycles or C value) were optimized for best accuracy by trial-and-error, and the classifier version that delivered the best results was selected and further validated with the external testing set (Fig. 1).

The training set overall included $n = 2300$ spectra, originating from $n = 88$ strains, from which 84 isolates were measured in Bremen (on CBA, CHO, TSA, MHA, MacC, SSA and XLD), and $n = 4$ *S. Typhi* isolates were measured in Bologna (only on CBA and TSA-SB). These strains corresponded to a total of 25 serogroups and 53 serovars. Each serogroup was represented by at least one isolate. The number of strains included for each serogroup varied in relation to their prevalence in the whole dataset (which reflects their frequency of isolation). Nevertheless, to minimize the unbalancing between the most frequent serogroups and the rarer ones, as well as to further test the robustness of the method, for the most numerous serogroups (O:4, O:7, O:8, O:9, O:3,10 and O:13), only the most common serovars were included. For each serovar, only one isolate was included (randomly chosen), except for the most numerous ones (*S.* Typhimurium, *S.* Enteritidis, etc.), for which the selection of the training isolates considered also the inter-serovar heterogeneity, previously investigated by PCA/LDA. Three isolates were included also for *S.* Senftenberg, given the high similarity degree between O:1,3,19 and O:3,10 groups and the need to maximize the differentiation capabilities for them.

The testing set included the remaining isolates measured in Bremen, grown on all culture media ($n = 468$) and the isolates measured at the external sites, grown on the culture media in use in the local routine workflow ($n = 327$ in Bologna, $n = 39$ in Japan and $n = 36$ in Torino). The composition of training and testing sets is shown in Table 1. The performance of the classifier was evaluated in terms of accuracy, error rate and failed classification rate, calculated for the "whole" isolate (comprising all spectra from all culture media) and also for the single media. Accuracy was defined as number of isolates correctly classified (green and yellow) out of the total number of isolates. Error rate was defined as number of isolates erroneously classified (misclassification, green and yellow) out of the total number of isolates. Failed classification rate was defined as number of isolates delivering a "red" result out of the total number of isolates. In cases where IRBT and reference method disagreed, IRBT analysis and agglutination test were repeated, to ensure that no mix-ups occurred during the analytical processes.

3. Results

Overall, 19,367 spectra were included in this study (16,515 measured in Bremen by 4 different operators, 2852 at external sites).

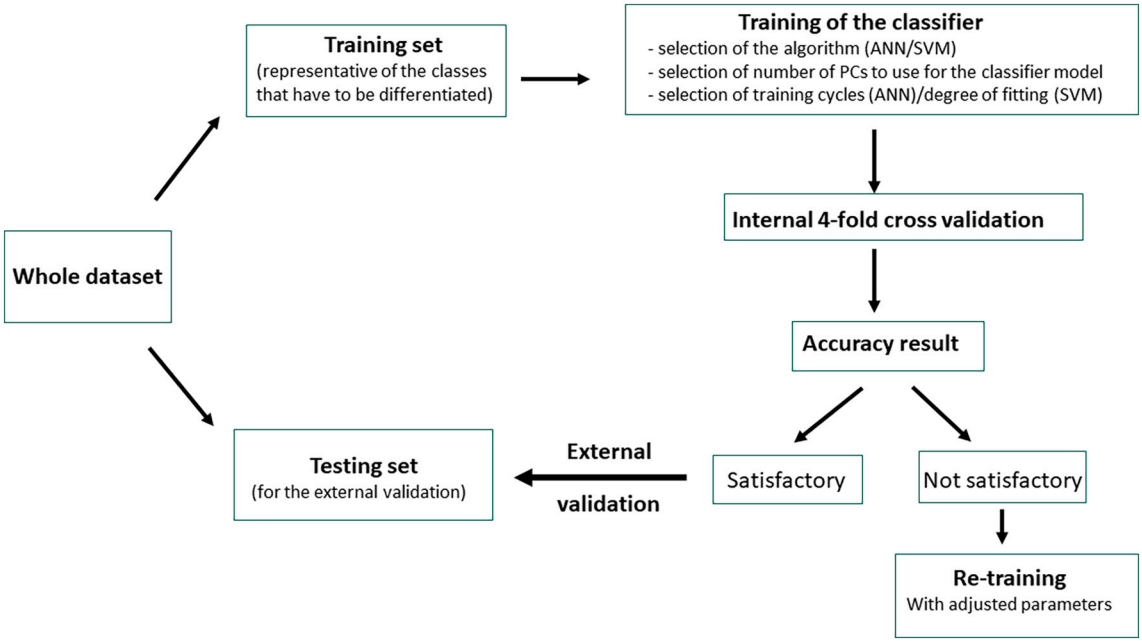

Fig. 1. Diagram of the IRBT general machine learning process.

Table 1
Accuracy (defined as number of isolates correctly classified in relation to the total number of isolates) and error rate (defined as number of isolates wrongly classified in relation to the total number of isolates) provided by the classifier on the different culture media. ChromID®, RAPID' Salmonella agar and TSA-SB were not included in the training set.

Media tested (nr. of isolates)	Center/s	Accuracy (%, n)	Error rate (%, n)
CBA (504)	Bremen, Torino	97.8 (493)	2.2 (11)
TSA (468)	Bremen	97.4 (456)	2.6 (12)
CHO (468)	Bremen	97.0 (454)	3.0 (14)
MHA (507)	Bremen, Okawa	97.2 (493)	2.8 (14)
MacC (468)	Bremen	96.4 (451)	3.6 (17)
XLD (468)	Bremen	95.1 (445)	4.9 (23)
SSA (468)	Bremen	94.7 (443)	5.3 (25)
ChromID®* (152)	Bremen	82.2 (125)	17.8 (27)
RAPID' Salmonella agar* (385)	Bremen	9.1 (35)	90.9 (350)
TSA-SB* (327)	Bologna	99.4 (325)	0.9 (2)

* Not included in the training set

3.1. Exploratory multivariate analysis

PCA and LDA showed good clustering of the different serogroups, with a clear differentiation of the vast majority of serogroups (Fig. 2).

A high degree of relatedness (i.e., a low spectral distance) was observed, as expected, between closely related serogroups (O:7/O:8 and O:3,10/O:1,3,19), but also between O:7/O:13/O:21. Nevertheless, the groups were clearly differentiable. A partial spectral overlapping was observed only for the non-paratyphoidal O:2 serovars (Nitra, Kiel and Koessen), some O:7 strains and the O:6,14,24 (*S.* Carrau) isolates.

Among the most numerous serovars included in this study, *S.* Typhimurium, *S.* Paratyphi B, *S.* Paratyphi C and *S.* Enteritidis showed a certain degree of heterogeneity in their distribution in the spectral space. On the contrary, *S.* Paratyphi A, *S.* Typhi, *S.* Infantis, *S.* Kentucky, *S.* Dublin, *S.* Goldcoast, *S.* Brandenburg and *S.* Derby showed a low heterogeneity. The differentiation among (para)typhoidal and non-(para)typhoidal serovars among serogroups O:2, O:4, O:7 and O:9 previously described by our group (Cordovana et al., 2021) could be confirmed. Nevertheless, no interference of this intra-serogroup differentiation could be detected with respect to the classification at O-serogroup level.

Concerning the different media, exploratory analysis showed a very high similarity between spectra measured from CBA, TSA, CHO, and MHA. Spectra from MacC and ChromID® Agar showed a lower similarity, but still often falling close within the spectral area. On the contrary, spectra from SSA, XLD and RAPID' Salmonella agar fell far away both from spectra measured from the non-selective media, and from each other (Fig. 3).

3.2. Performance of the automated classifiers

The classifier version showing the best results in the cross-validation was built with the linear SVM algorithm using 20 PCs, and its performance was further tested with external testing sets. All isolates delivered a classification result with an outlier value ≤4.0. The outlier was not used as an indicator of the reliability of the classification since no reasonable threshold values could exclude misclassifications. This is likely because the training set contains a high diversity of serogroups and many different media. It therefore spans a wide spectral space, rendering exclusion of outliers by a measure of distance futile.

Among the isolates tested in Bremen, accuracy for the different culture media included in the training set was 97.6% (457/468) for CBA, 97.0% (454/468) for CHO, 97.4% (456/468) for TSA, 97.7% (458/468) for MHA, 96.4% (451/468) for MacC, 95.1% (445/468) for XLD and 94.7% (443/468) for SSA. Considering the two chromogenic media (not

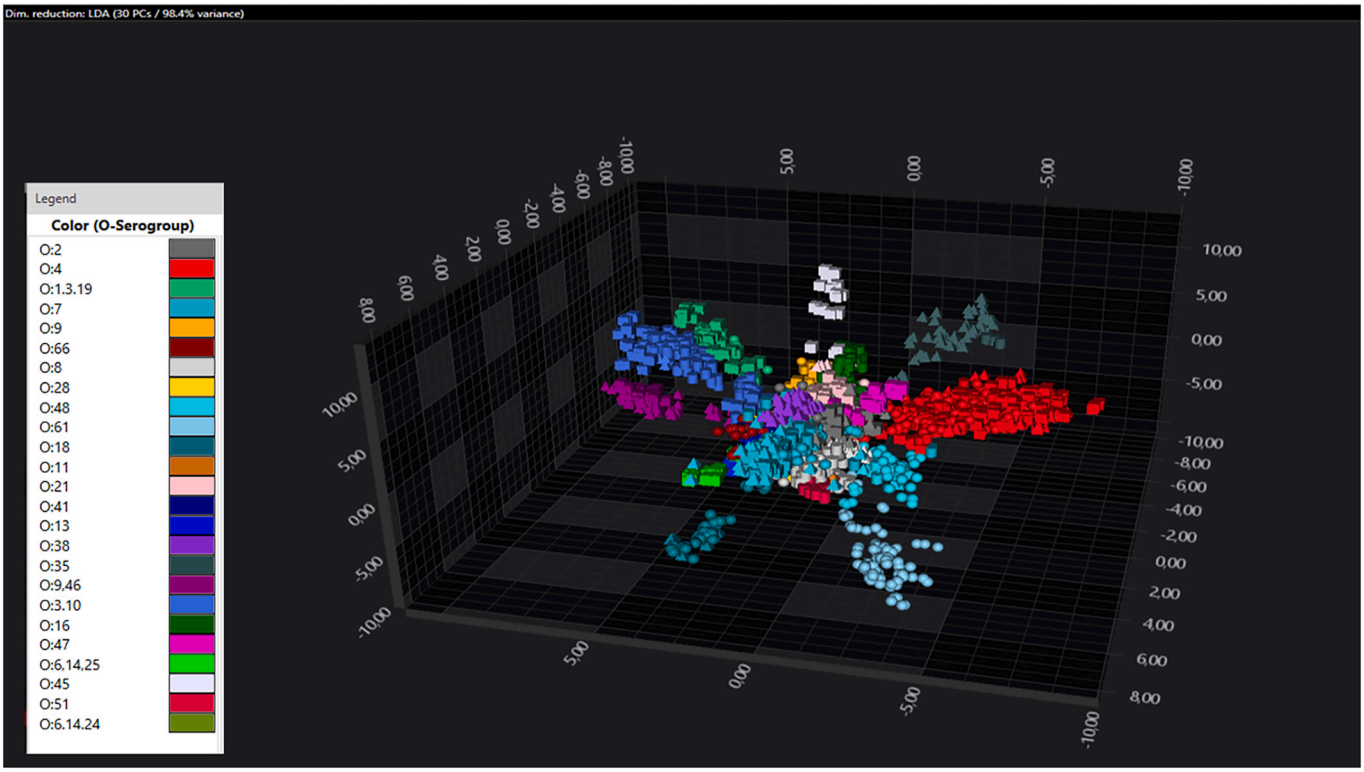

Fig. 2. 3D scatter plot showing the distribution of O-serogroups of *Salmonella* species in IR spectral space. Thirty principal components were used to create an LDA by O groups. The first three LD axes are shown in the diagram. The spectra are colored by O-serogroups, and the shapes correspond to different isolates. The discriminability of so many groups can be achieved in multidimensional space, but will not be visible in a low-dimensional scatter plot.

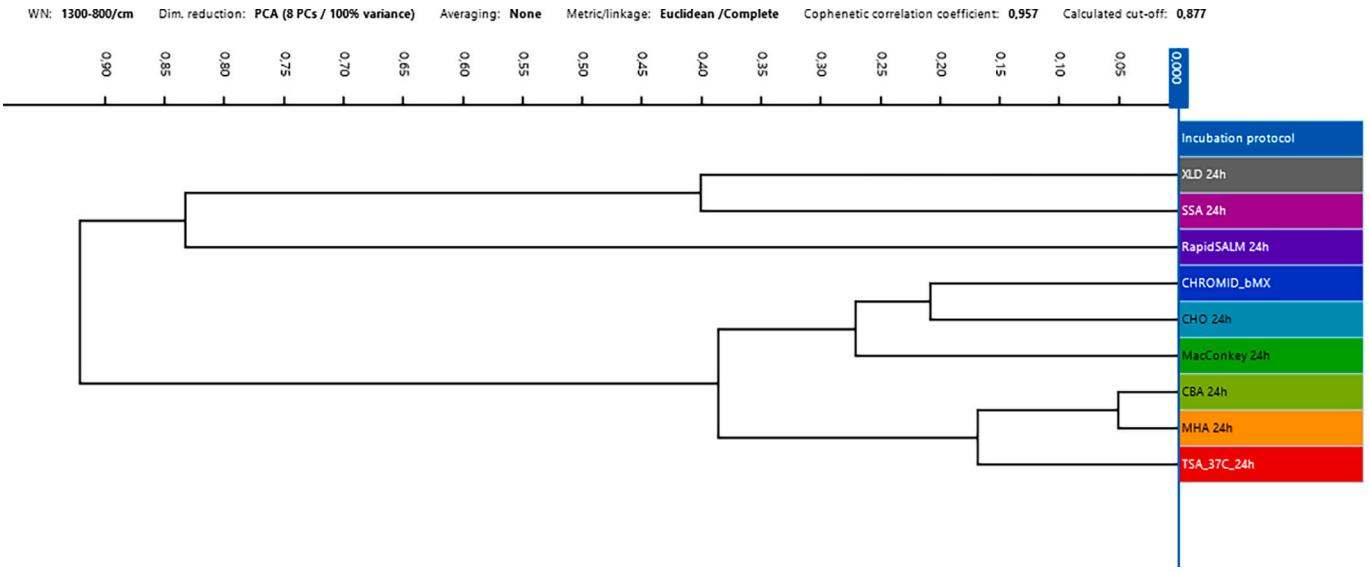

Fig. 3. Hierarchical cluster analysis of spectra of one isolate measured from different media (listed in the column on the right side of the image – "Incubation protocol"), performed using Euclidean average linkage. On the left side of the figure, the dendrogram is displayed, showing that the non-selective media form one cluster, together with MacC and ChromID® agar, while the iron-containing media (XLD and SSA) and RAPID Agar form another more diverse cluster.

included in the training set), accuracy was 82.2 (125/152) for ChromID® Agar and 9.1 (35/385), for RAPID' Salmonella agar.

Overall, 451/468 (96.4%) isolates were correctly identified from all non-selective media. Among them, 427/468 (91.3%) were correctly classified also from selective media, while 24/468 (5.1%) showed a misclassification result for one selective medium (($n = 11$ from XLD, n = 11 from SSA and $n = 2$ from MacC). Another eight isolates (1.7%) showed a misclassification result for one or more media, among which a non-selective one (n = 1 *S.* Strathcona, n = 1 *S.* Goldcoast, n = 1 *S.* Infantis, n = 1 *S.* Panama, n = 1 *S.* Enteritidis, n = 1 *S.* Paratyphi C, n = 1 *S. enterica* subsp. enterica O:9, n = 1 *S.* diarizonae O:48). Nine isolates (1.9%) were wrongly classified from all media. These latter nine isolates comprise four strains exhibiting a rough phenotype (n = 2 *S.* Typhimurium and n = 2 *S.* Enteritidis), n = 2 *S.* diarizonae (O:11 and O:7), n = 1 *S.* Senftenberg, n = 1 *S.* Isangi and n = 1 *S.* Anatum. Agglutination testing confirmed the original typing result.

Among the isolates from Okawa, 35/39 (89.8%) were correctly classified. The misclassifications involved n = 1 *S.* Kande (O:1,3,19), n = 1 *S.* Memphis (O:18), n = 1 *S.* Harburg (O:6,14) and n = 1 *S.* Claibornei (O:9).

Among the isolates from Bologna, 324/327 (99.1%) were correctly classified. Misclassification involved n = 2 O:16 isolates (n = 1 *S.* Szentes and n = 1 *S.* Hvittingfoss).

Among the isolates from Torino, 36/36 (100%) were correctly classified.

A summary of the results is shown in Table 1.

Accuracy for the different serogroups is reported in Table 2. For strains measured in Bremen, isolates were considered and counted as correctly classified only in those cases in which they showed a correct classification result from all the seven media included in the training set.

Table 2
Accuracy delivered by the IRBT classifier stratified for O-serogroups. Regarding the O:2 serogroup, accuracy for *S. paratyphi* A was 100% (19/19), as the misclassification results involved only non-Paratyphi A serovars (*S.* Kiel and *S.* Koessen).

O-serogroup	Total. isolates in the testing set (n)	Accuracy (%, n)
O:4	398	97.5 (388)
O:9	180	95.6 (172)
O:7	106	94.4 (100)
O:8	54	96.3 (52)
O:3,10	29	96.6 (28)
O:13	24	87.5 (21)
O:2	23	91.3 (21)
O:11	16	100 (16)
O:28	10	90.0 (9)
O:1,3,19	8	87.5 (7)
O:16	4	100 (4)
O:18	4	100 (4)
O:48	3	66.7 (2)
O:61	3	100 (3)
O:38	2	100 (2)
O:9,46	2	100 (2)
O:21	1	100 (1)
O:35	1	100 (1)
O:47	1	100 (1)
O:6,14,25	1	0 (0)
O:41	–	–
O:6,14,24	–	–
O:45	–	–
O:51	–	–
O:66	–	–
Tot.	870	95.9 (834)

4. Discussion

Detection and appropriate identification of *Salmonella enterica* isolates is of importance for various purposes, ranging from clinical microbiology to veterinary medicine, food hygiene and environmental monitoring. Different levels of typing are required for the different purposes, but all of them are technically challenging because of the required deepness of intra-species discrimination involved (serogroup, serotype, strain type in case of outbreaks) and comprise both expensive and laborious analytical methods. From the clinical point of view, recognition of typhoidal serovars is crucial to estimate their etiological relevance, to optimize the calculated treatment and to recognize potential outbreak situations. Similarly, the recognition of specific serogroups/serotypes is pivotal in for food and veterinary hygiene to draw properly epidemiological layouts.

FTIR spectroscopy is a technology explored for typing of different bacterial genera and species (especially Enterobacterales, *Enterococcus*, *Streptococcus*, *Staphylococcus, Listeria, Bacillus* and *Lactobacillus* species) (Quintelas et al., 2018). It has been proposed for, typing below species level, including the detection of multidrug resistant bacteria outbreaks, aiming at defining an alternative to the established, but cost-intensive,

time-consuming and still hardly large-scale applicable DNA-based techniques commonly used for epidemiological purposes (Sabat et al., 2013; Dinkelacker et al., 2018; Burckhardt et al., 2019; Martak et al., 2019; Vogt et al., 2019; Hu et al., 2021; Guerrero-Lozano et al., 2022; Lombardo et al., 2021)In this study, we evaluated the performance of the FTIR-based IRBT system to discriminate *Salmonella* serogroups by testing >900 well characterized isolates, including 25 serogroups and 138 serovars, applying machine learning. The strain collection comprised the globally most frequently isolated serovars (Enteritidis, Infantis, Typhimurium and its monophasic variant) but also the clinically most relevant ones (Typhi, Paratyphi), the emerging ones (Rissen, Heidelberg, Infantis, Newport, Mbandaka, Stanley, Senftenberg) and several less common ones (Table 1). The strains were collected in several countries, from different origins (human, animal, food, environment, culture collections), in some cases prospectively, in other cases selected based on precise requirements (paratyphoidal serovars, rare serogroups, etc.), in order to include the broadest biological, geographic and epidemiological variability possible in this multicentric evaluation.

In the first part of the study, an exploratory analysis with PCA/LDA was performed to get an overview on the discriminatory power between the O-serogroups, and to spot possible issues and sources of misclassification. Each serogroup exhibits a high spectral variance, related to 1) the presence of 25 serogroups and >100 serovars and 2) the inclusion of spectra from isolates grown on very heterogeneous culture media. Notwithstanding, the exploratory analysis with PCA/LDA enabled a clear differentiation of the most serogroups, with a spectral distance coherent with their genomic and antigenic relatedness. When looking at the 3D scatterplot, an apparent and partial overlapping of serogroups could be observed in the spectral space where the serogroups O:2/O:21/O:4/O:13/O:11 and O:6,14,24 are located. Nevertheless, the 3D scatterplot shows only the first three dimensions, while the total number of dimensions included in this LDA analysis is 24 (the total number of dimensions in an LDA analysis is, by definition, equal to the number of classes (O-serogroups) minus one). Despite the fact that often the first 3 dimensions are the most discriminating, in several cases the discrimination of some classes can be achieved in dimensions beyond the third, and in those cases it is not visible in a 3D scatter plot. In this study, the machine learning algorithms, that takes into account all the dimensions, proved to be able to differentiate also the above-mentioned serogroups with high accuracy,(Table 2).In concordance with the previous study by Campos et al. (Campos et al., 2018), the discrimination of closely related serogroups such as O:3,10 and O:1,3,19 (former E_1-E_2-E_3 and E_4), O:7 and O:8 (formerly C_1 and C_2-C_3) and O:9 and O:9,46 (former D_1 and D_2) was also reliably achieved.

In the last years, several studies evaluated the potential of machine learning applied in general to the field of *Salmonella* (Tanui et al., 2022; Bolinger et al., 2021; Munck et al., 2020; Nguyen et al., 2019). Nevertheless, all of them mainly focused on epidemiological (source tracking) or genetic purpose, and none investigated this approach for typing at infra-species level. In this study, the novel artificial intelligence capabilities implemented into the IRBT software were thoroughly investigated, to develop a classifier that allows the differentiation of *Salmonella* isolates at O-serogroup level. The training set included 25 serogroups, 53 serovars and seven among the globally most widely used solid agar-based culture media. Different machine learning algorithms, set with different training parameters, were applied and the classifier which showed the best accuracy in the internal cross-validation (build with linear SVM) was then tested with external datasets. To assess the robustness of the method, the validation set included isolates measured in the Bruker laboratory, as well as isolates measured at different sites by collaboration partners. The classifier delivered a good accuracy, with all datasets. For the strains measured in Bremen accuracy was >97% for non-selective media, and ranged from 94.7 to 96.4% for selective media). Overall, 427 out of 468 isolates (91.3%) were correctly classified from all media, while further 32 (6.8%) showed a misclassification for one or two media (in most cases SSA or XLD agar). These findings show that the considerably different growth conditions resulting from the use of selective media did not have a relevant impact on the typing capabilities of the IRBT software, which enabled a data analysis that succeeded in nullifying the spectral differences due to different incubation conditions. Also, the presence of the black precipitate which *Salmonella* isolates can produce in iron-containing media did not seem to interfere significantly with the IRBT typing.

Nine isolates were incorrectly classified from any culture medium. In all cases, the PCA/LDA analysis was coherent with the result delivered by machine learning, as the spectra of those isolates were located far away from the serogroup they belong to. Possible explanation for this misclassification could be an atypical phenotype (four isolates grew with rough colonies), or other unusual strain-related features (two isolates belonged to non-enterica subspecies).

Surprisingly, the classifier delivered a moderate accuracy even with the testing set of isolates grown on ChromID® Agar (82.2%), The accuracy with another chromogenic medium, RAPID' Salmonella agar, was very poor (9%), so the use of chromogenic media for FTIR spectroscopy should be evaluated on case-by-case basis, considering the possibility to build a specific classifier.

5. Conclusion

In this study, the IRBT system proved to be a promising and useful tool for *Salmonella* typing at the O-serogroup level. The application of artificial intelligence enables a novel approach, which is fully automated, and does not need any operator-depending interpretation of the results. The classifier can be applied on already acquired spectra, or can be implemented into the IRBT measurement software, allowing the classification in real time during the spectra acquisition. The simple sample preparation, the short handling time, the user-friendliness of the software, the possibility to analyze samples as soon as colonies on standard solid agar media are grown, as well as the system's lacking requirement of a minimum number of samples to set up a run or to optimize reagents, should enable an easy implementation of FTIR spectroscopy into routine laboratories as a rapid typing technology. In contrast to reference methods, which are mainly used retrospectively, IRBT can be used for real-time investigation providing results in a very short time (20 min −2 h) after cultivation of bacteria.

Further studies including more serogroups, more isolates of the rarest serogroups and more spectra from selective media will be necessary to widen the typing capabilities, as well as to further strengthen the robustness of the method. In addition, an evaluation of the potential of the method in terms of classification of serovars or even at the strain level would be desirable, in order to define the usefulness of the system and also for the tracking and monitoring of outbreaks.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial or not-for-profit sectors.

Author contributions

Conceptualization M.C., N.M., O.J.L., M.P., S.Z., M.K.; methodology M.C., N.M.; software N.M.; validation: M.C., N.M., O.J.L., M.P., S.Z.; formal analysis: M.C., N.M.; investigation: M.C., M.A., S.L., M.P., S.Z., M. P., H.M.H., A.L., L.S.; resources: M.B., J.St., H.F., D.D., Y.F., Z.N., J.So., L.O., A.C.V., U.S.J., S.A., S.L., A.W., S.R., R.M.H., J.M., A.B.P.; data curation: M.C., N.M.; Writing – original draft: M.C., H.F., D.D., O.J.L.; Writing – review and editing: N.M., F.G., S.L., M.P., S.Z., M.B., J.St., H. F., D.D.; Y.F., Z.N., J. So., L.O., U.S.J., H.M.H., A.L., S.P., L.S., A,W., S.R., R.M.H., J.M., A.B.P., M.K.; visualization: M.C., N.M., O.J.L., supervision: N.M., M.K., H.F., D.D., O.J.L.; project administration: M.C.; funding acquisition: none.

Aknowledgment

The authors would like to thank Susanne Hecht, Beatrix Wegemann and Leon Gebauer for the laboratory support.

Declaration of Competing Interest

O.J.L, F.G., S.L., M.A., M.P., S.Z., M.B., J.S., H.F., D.D., Y.F., Z.N., J. S., L.O., A.C.V., U.S.J., H.M.H., A.L., S.A., S.P., L.S., A.W., S.R., R.M.H., J.M. and A.B.P. declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest. M.C., N.M. and M.K. are Bruker's employees.

Data availability

Data will be made available on request.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.mimet.2022.106564.

References

Antillón, M., Warren, J.L., Crawford, F.W., Weinberger, D.M., Kürüm, E., Pak, G.D., Marks, F., Pitzer, V.E., 2017. The burden of typhoid fever in low- and middle-income countries: a meta-regression approach. PLoS Negl. Trop. Dis. 11 (2), e0005376 https://doi.org/10.1371/journal.pntd.0005376.

Antunes, P., Mourão, J., Campos, J., Peixe, L., 2016. Salmonellosis: the role of poultry meat. Clin. Microbiol. Infect. 22, 110–121. https://doi.org/10.1016/j.cmi.2015.12.004.

Antunes, P., Campos, J., Mourão, J., Ribeiro, T., Novais, C., Peixe, L., 2017. High occurrence and unusual serotype diversity of non-typhoidal *Salmonella* in non-clinical niches. Angola. Epidemiol. Infect. 145, 883–886. https://doi.org/10.1017/S095026881600296X.

Baker, M.J., Trevisan, J., Bassan, P., Bhargava, R., Butler, H.J., Dorling, K.M., Fielden, P. R., Fogarty, S.W., Fullwood, N.J., Heys, K.A., Hughes, C., Lasch, P., Martin-Hirsch, P. L., Obinaju, B., Sockalingum, G.D., Sulé-Suso, J., Strong, R.J., Walsh, M.J., Wood, B. R., Gardner, P., Martin, F.L., 2014. Using Fourier transform IR spectroscopy to analyze biological materials. Nat. Protoc. 9, 1771–1791. https://doi.org/10.1038/nprot.2014.110.

Barco, L., Lettini, A.A., Ramon, E., Longo, A., Saccardin, C., Dalla Pozza, M.C., Ricci, A., 2011. A rapid and sensitive method to identify and differentiate *Salmonella enterica* serotype typhimurium and Salmonella enterica serotype 4,[5],12:i:- by combining traditional serotyping and multiplex polymerase chain reaction. Foodborne Pathog. Dis. Jun 8 (6), 741–743. https://doi.org/10.1089/fpd.2010.0776.

Bolinger, H., Tran, D., Harary, K., Paoli, G.C., Guron, G.K.P., Namazi, H., Khaksar, R., 2021. Utilizing the microbiota and machine learning algorithms to assess risk of Salmonella contamination in poultry Rinsate. J. Food Prot. 84 (9), 1648–1657. https://doi.org/10.4315/JFP-20-367. Sep 1.

Burckhardt, I., Sebastian, K., Mauder, N., Kostrzewa, M., Burckhardt, F., Zimmermann, S., 2019. Analysis of *Streptococcus pneumoniae* using Fourier-transformed infrared spectroscopy allows prediction of capsular serotype. Eur. J. Clin. Microbiol. Infect. Dis. 38 (10), 1883–1890. https://doi.org/10.1007/s10096-019-03622-y.

Campos, J., Sousa, C., Mourão, J., Lopez, 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.

CDC (Centers for Disease Control and Prevention), 2017. Foodborne Diseases Active Surveillance Network (FoodNet). Foodnet 2015 Surveillance Report (Final Data). Atlanta, Georgia. https://www.cdc.gov/foodnet/pdfs/FoodNet-Annual-Report-2015-508c.pdf.

Cordovana, M., Mauder, N., Kostrzewa, M., Wille, A., Rojak, S., Hagen, R.M., Ambretti, S., Pongolini, S., Soliani, L., Justesen, U.S., Holt, H.M., Join-Lambert, O., Lehello, S., Auzou, M., Veloo, A.C., 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 (4), 853. https://doi.org/10.3390/microorganisms9040853, 15.

Davis, R., Mauer, L.J., 2010. Fourier transform infrared (FTIR) spectroscopy: a rapid tool for detection and analysis of foodborne pathogenic bacteria. Curr Res. Technol. Educ. Topics Appl. Microbiol. Biotechnol. 2, 1582–1594.

De Lamo-Castellví, S., Männing, A., Rodríguez-Saona, L.E., 2010. Fourier-transform infrared spectroscopy combined with immunomagnetic separation as a tool to discriminate *Salmonella* Serovars. Analyst. 135 (11), 2987–2992. https://doi.org/10.1039/c0an00497a.

Dinkelacker, A., Vogt, S., Oberhettinger, P., Mauder, N., Rau, J., Kostrzewa, M., Rossen, J.W.A., Autenrieth, I.B., Peter, S., Liese, J., 2018. Typing and species identification of clinical *Klebsiella* isolates by Fourier transform infrared spectroscopy and matrix-assisted laser desorption ionization-time of flight mass spectrometry. J. Clin. Microbiol. https://doi.org/10.1128/JCM.00843-18, 56e00843-18.

EFSA, 2010. Scientific Opinion on monitoring and assessment of the public health risk of "*Salmonella Typhimurium*-like" strains. EFSA Panel on Biological Hazards (BIOHAZ). EFSA J. 8 (10), 1826 and appendix, 1–41. https://doi.org/10.2903/j.efsa.2010.1826.

EFSA and ECDC (European Food Safety Authority and European Centre for Disease Prevention and Control), 2017. The European union summary report on trends and sources of zoonoses, zoonotic agents and food-borne outbreaks in 2016. EFSA J. 15, 5077. https://ecdc.europa.eu/sites/portal/files/documents/summary-report-zoonoses-foodborne-outbreaks-2016.pdf.

Griffiths, P.R., De Haseh, J.A., 2007. Fourier Transform Infrared Spectrometry. John Wiley & Sons, Inc., Hoboken, New Jersey.

Grimont, P.A., Weill, F.X., 2007. Antigenic Formulae of the *Salmonella* Serovars. WHO Collaborating Center for Reference and Research on Salmonella, Paris, France.

Guerrero-Lozano, I., Galán-Sánchez, F., Rodríguez-Iglesias, M., 2022. Fourier transform infrared spectroscopy as a new tool for surveillance in local stewardship antimicrobial program: a retrospective study in a nosocomial *Acinetobacter baumannii* outbreak. Braz. J. Microbiol. 30 https://doi.org/10.1007/s42770-022-00774-6.

Helm, D., Labischinski, H., Schallehn, G., Naumann, D., 1991. Classification and identification of bacteria by Fourier-transform infrared spectroscopy. J. Gen. Microbiol. 137, 69–79. https://doi.org/10.1099/00221287-137-1-6982.

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. 4, 1343–1352. https://doi.org/10.1111/1751-7915.13709.

Kim, S., Kim, H., Reuhs, B.L., Mauer, L.J., 2006. Differentiation of outer membrane proteins from *Salmonella enterica* serotypes using Fourier transform infrared spectroscopy and chemometrics. Lett. Appl. Microbiol. 42 (3), 229–234. https://doi.org/10.1111/j.1472-765X.2005.01828.x.

Lasch, P., Naumann, D., 2015. Infrared spectroscopy in microbiology. In: Encyclopedia of Analytical Chemistry. John Wiley & Sons, Ltd, Chichester, pp. 1–32.

Lombardo, D., Cordovana, M., Deidda, F., Pane, M., Ambretti, S., 2021. Application of Fourier transform infrared spectroscopy for real-time typing of *Acinetobacter baumannii* outbreak in intensive care unit. Future Microbiol. 16, 1239–1250. https://doi.org/10.2217/fmb-2020-0276.

Männing, A., Baldauf, N.A., Rodriguez-Romo, L.A., Yousef, A.E., Rodríguez-Saona, L.E., 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 (11), 2249–2256. https://doi.org/10.4315/0362-028x-71.11.2249.

Martak, D., Valot, B., Sauget, M., Cholley, P., Thouverez, M., Bretrand, X., Hocquet, D., 2019. Fourier-transform infrared spectroscopy can quickly type gram-negative bacilli responsible for hospital outbreaks. Front. Microbiol. 10, 1440. https://doi.org/10.3389/fmicb.2019.01440.

Mourão, J., Machado, J., Novais, C., Antunes, P., Peixe, L., 2014. Characterization of the emerging clinically-relevant multidrug-resistant *Salmonella enterica* serotype 4,[5],12:i:- (monophasic variant of *S. typhimurium*) clones. Eur. J. Clin. Microbiol. Infect. Dis. 15, 15. https://doi.org/10.1007/s10096-014-2180-1.

Munck, N., Njage, P.M.K., Leekitcharoenphon, P., Litrup, E., Hald, T., 2020. Application of whole-genome sequences and machine learning in source attribution of Salmonella typhimurium. Risk Anal. Sep 40 (9), 1693–1705. https://doi.org/10.1111/risa.13510.

Naumann, D., Helm, D., Labischinski, H., 1991. Microbiological characterization by FT-IR spectroscopy. Nature 351, 81–82. https://doi.org/10.1039/b418288j.

Nguyen, M., Long, S.W., McDermott, P.F., Olsen, R.J., Olson, R., Stevens, R.L., Tyson, G. H., Zhao, S., Davis, J.J., 2019. Using machine learning to predict antimicrobial MICs and associated genomic features for Nontyphoidal *Salmonella*. J. Clin. Microbiol. 57 (2) https://doi.org/10.1128/JCM.01260-18, 30. e01260-18.

Painter, J.A., Hoekstra, R.M., Ayers, T., Tauxe, R.V., Braden, C.R., Angulo, F.J., Griffin, P. M., 2013. Attribution of foodborne illnesses, hospitalization, and deaths to food commodities by using outbreak data, United States, 1998-2008. Emerg. Infect. Dis. 19 (3), 407–415. https://doi.org/10.3201/eid1903.111866.

Parmley, E.J., Pintar, K., Majowicz, S., Avery, B., Cook, A., Jokinen, C., Gannon, V., Lapen, D.R., Topp, E., Edge, T.A., Gilmour, M., Pollari, F., Reid-Smith, R., Irwin, R., 2013. A Canadian application of one health: integration of *Salmonella* data from various Canadian surveillance programs (2005-2010). Foodborne Pathog. Dis. 10 (9), 747–756. https://doi.org/10.1089/fpd.2012.1438.

Preisner, O., Lopes, J.A., Guiomar, R., Machado, J., Menezes, J.C., 2007. Fourier transform infrared (FT-IR) spectroscopy in bacteriology: towards a reference method for bacterial discrimination. Anal. Bioanal. Chem. 387, 1739–1748. https://doi.org/10.1007/s00216-006-0851-1.

Preisner, O., Guiomar, R., Machado, J., Menezes, J.C., Lopes, J.A., 2010. Application of Fourier transform infrared spectroscopy and chemometrics for differentiation of *Salmonella enterica* serovar Enteritidis phage types. Appl. Environ. Microbiol. 76, 3538–3544. https://doi.org/10.1128/AEM.01589-09.

Quintelas, C., Ferreira, E.C., Lopes, J.A., Sousa, C., 2018. An overview of the evolution of infrared spectroscopy applied to bacterial typing. Biotechnol. J. 13, 1700449. https://doi.org/10.1002/biot.201700449.

Ribot, E.M., Fair, M.A., Gautom, R., Cameron, D.N., Hunter, S.B., Swaminathan, B., Barret, T.J., 2006. Standardization of pulsed-field gel electrophoresis protocols for

the subtyping of *Escherichia coli* O157:H7, Salmonella, and Shigella for PulseNet. Foodborne Pathog Dis. Spring 3 (1), 59–67. https://doi.org/10.1089/fpd.2006.3.59 (PMID: 16602980).

Sabat, A.J., Budimir, A., Nashev, D., Sá-Leão, R., van Dijl, J.M., Laurent, F., Grundmann, H., Friedrich, A.W., ESCMID Study Group of Epidemiological Markers (ESGEM), 2013. Overview of molecular typing methods for outbreak detection and epidemiological surveillance. Euro Surveill. 18, 20380. https://doi.org/10.2807/ese.18.04.20380-en.

Stuart, B.H., 2004. Infrared Spectroscopy: Fundamentals and Applications. John Wiley and Sons Ltd., Chichester, United Kingdom.

Sundaram, J., Park, B., Hinton Jr., A., Yoon, S.C., Windham, W.R., Lawrence, K.C., 2012. Classification and structural analysis of live and dead *Salmonella* cells using Fourier transform infrared spectroscopy and principal component analysis. J. Agric. Food Chem. 60 (4), 991–1004. https://doi.org/10.1021/jf204081g, 1.

Tang, S., Orsi, R.H., Luo, H., Ge, C., Zhang, G., Baker, R.C., Stevenson, A., Wiedmann, M., 2019. Assessment and comparison of molecular subtyping and characterization methods for *Salmonella*. Front. Microbiol. 12 (10), 1591. https://doi.org/10.3389/fmicb.2019.01591.

Tanui, C.K., Karanth, S., Niage, P.M.K., Meng, J., Pradhan, A.K., 2022. Machine learning-based predictive modeling to identify genotypic traits associated with Salmonella enterica disease endpoints in isolates from ground chicken. LWT 154, 112701. https://doi.org/10.1016/j.lwt.2021.112701.

Tennant, S.M., Diallo, S., Haim Levy, H., Livio, S., Sow, S.O., Tapia, M., Fields, P.I., Mikoleit, M., Tamboura, B., Kotloff, K.L., Nataro, J.P., Galen, J.E., Levine, M.M., 2010. Identification by PCR of non-typhoidal *Salmonella enterica* serovars associated with invasive infections among febrile patients in Mali. PLosNegl. Trop. Dis. 4, e621 https://doi.org/10.1371/journal.pntd.0000621.

Vogt, S., Löffler, K., Dinkelacker, A.G., Bader, B., Autenrieth, I.B., Peter, S., Liese, J., 2019. Fourier-transform infrared (FTIR) spectroscopy for typing of clinical *Enterobacter cloacae* complex isolates. Front. Microbiol. 10, 2582. https://doi.org/10.3389/fmicb.2019.02582.[21].

Wenning, M., Scherer, S., 2013. Identification of microorganisms by FTIR spectroscopy: perspective and limitations of the method. Appl. Microbiol. Biotechnol. 97, 7111–7120. https://doi.org/10.1007/s00253-013-5087-3.

World Health Organization, 2016. WHO Estimates of the Global Burden of Foodborne Diseases. Foodborne Disease Burden Epidemiology Reference Group, pp. 2007–2015. http://www.who.int/foodsafety/areas_work/foodborne-diseases/ferg/en/.

Zamowiec, P., Lechowicz, L., Czerwonka, G., Kaca, W., 2015. Fourier transform infrared spectroscopy (FTIR) as a tool for the identification and differentiation of pathogenic bacteria. Curr. Med. Chem. 22, 1710–1718. https://doi.org/10.2174/0929867322666150311152800.