

Automated discrimination of *Salmonella* Typhi by Fourier-Transform Infrared (FTIR) spectroscopy and machine learning

Miriam Cordovana¹, Stefan Zimmermann², Tabea Gaube², Simone Ambretti^{3,4}, Stefano Pongolini⁵, Laura Soliani⁵, Javiera Jiménez-Rodríguez⁶, Patricia García⁶, Princess Morales⁷, Mohammed Suleiman^{7,8}, Andres Perez-Lopez^{7,9}, Olivier Join-Lambert¹⁰, Simon Le Hello¹⁰, Michel Auzou¹⁰, Denise Dekker¹¹, Arthur B. Pranada¹², Helene Oberreuter¹³, Joerg Rau¹³, Hagen Frickmann^{14,15}, Markus Kostrzewa¹

1) Bruker Daltonic GmbH, Bremen, Germany 2) University Hospital Heidelberg, Germany 3) Microbiology Unit, IRCCS Azienda Ospedaliero-Universitaria di Bologna, Italy 4) Microbiology Unit, DIMEC, Alma Mater Studiorum Università di Bologna, Italy 5) Risk Analysis and Genomic Epidemiology Unit, Istituto Zooprofilattico Sperimentale della Lombardia e dell'Emilia-Romagna, Parma Italy 6) Departamento de Laboratorios Clínicos, Escuela de Medicina, Facultad de Medicina, Pontificia Universidad Católica de Chile, Santiago, Chile 7) Department of Pathology, Sidra Medicine, P.O. Box 26999, Doha, Qatar 8) Department of Biomedical Sciences, Qatar University, P.O. Box 2713, Doha, Qatar 9) Department of Pathology and Laboratory Medicine, Weill Cornell Medicine in Qatar, P.O. Box 24144, Doha, Qatar 10) Université de Caen, Normandie, Caen, France 11) One Health Bacteriology working group, Bernhard Nocht Institute for Tropical Medicine Hamburg, Hamburg, Germany 12) MVZ Dr. Eberhard & Partner Dortmund, Department of Medical Microbiology, Dortmund, Germany 13) Chemical and Veterinary Analysis Agency (CVUA) Stuttgart – Fellbach, Germany 14) Department of Microbiology and Hospital Hygiene, Bundeswehr Hospital Hamburg, Hamburg, Germany 15) Institute for Medical Microbiology, Virology and Hygiene, University Medicine Rostock, Rostock, Germany

Background

Salmonella enterica serovar Typhi (S. Typhi) is the etiological agent of typhoid fever, a life-threatening illness. Spread via contaminated water and food, it remains a major cause of bacteraemia and sepsis, particularly in resource-limited countries. Showing partly comparable symptoms, typhoid fever is an important differential diagnosis to malaria in endemic areas.

S. Typhi is typically identified using a combination of classical microbiology, biochemical testing, serotyping, and increasingly molecular methods. Reliable and user-friendly alternatives to laborious and resource-demanding serological/genome-based methods are desirable. In a previous analysis, Fourier transform infrared spectroscopy (FTIR) proved to be able to differentiate S. Typhi from other serovars of the serogroup O:9 [1].

In the here-presented study, a **classifier for the automated discrimination of S. Typhi**, based on FTIR and machine learning, was established and tested.

Material and methods

✓ Sample set

✓ Overall, **N=355 *Salmonella enterica* isolates** belonging to the **O:9 serogroup** were included in this study:

- n=48 S. Typhi
- n=199 S. Enteritidis
- n=52 S. Dublin
- n=10 S. Napoli
- n=46 belonging to other serovars (Gallinarum, Panama, Javiana, Eastbourne, Zaiman, Goettingen, Kapemba, Berta)

✓ The isolates were collected from clinical, environmental and food samples, at different geographical sites.

✓ The dataset comprised two subsets:

- **Study design** dataset (n=193)
- **External independent testing** dataset (n=162)

FT-IR analysis

✓ FT-IR analysis was performed applying the IR Biotyper® system (IRBT - Bruker Daltonics, Germany). Spectra were acquired from bacterial suspensions in ethanol solution, using cultures on Columbia blood agar and tryptic soy agar with/without 5% sheep blood (different manufacturers) incubated overnight at 35 ° C, on three independent days.

✓ **Principal component analysis (PCA)** and **linear discriminant analysis (LDA)** were used for exploratory data analysis.

✓ **Machine learning** was applied to **build a classifier** to distinguish S. Typhi from S. non-Typhi, using the Support Vector Machine (SVM) algorithm included in the IR Biotyper® software. The classifier was trained with n=141 isolates (n=27 S. Typhi and n=114 S. non-Typhi) and tested with the remaining n=214 ones.

Results

➤ **Exploratory data analysis** applying PCA/LDA showed a clear separation between S. Typhi and the other O:9 serovars (**Figure 1**).

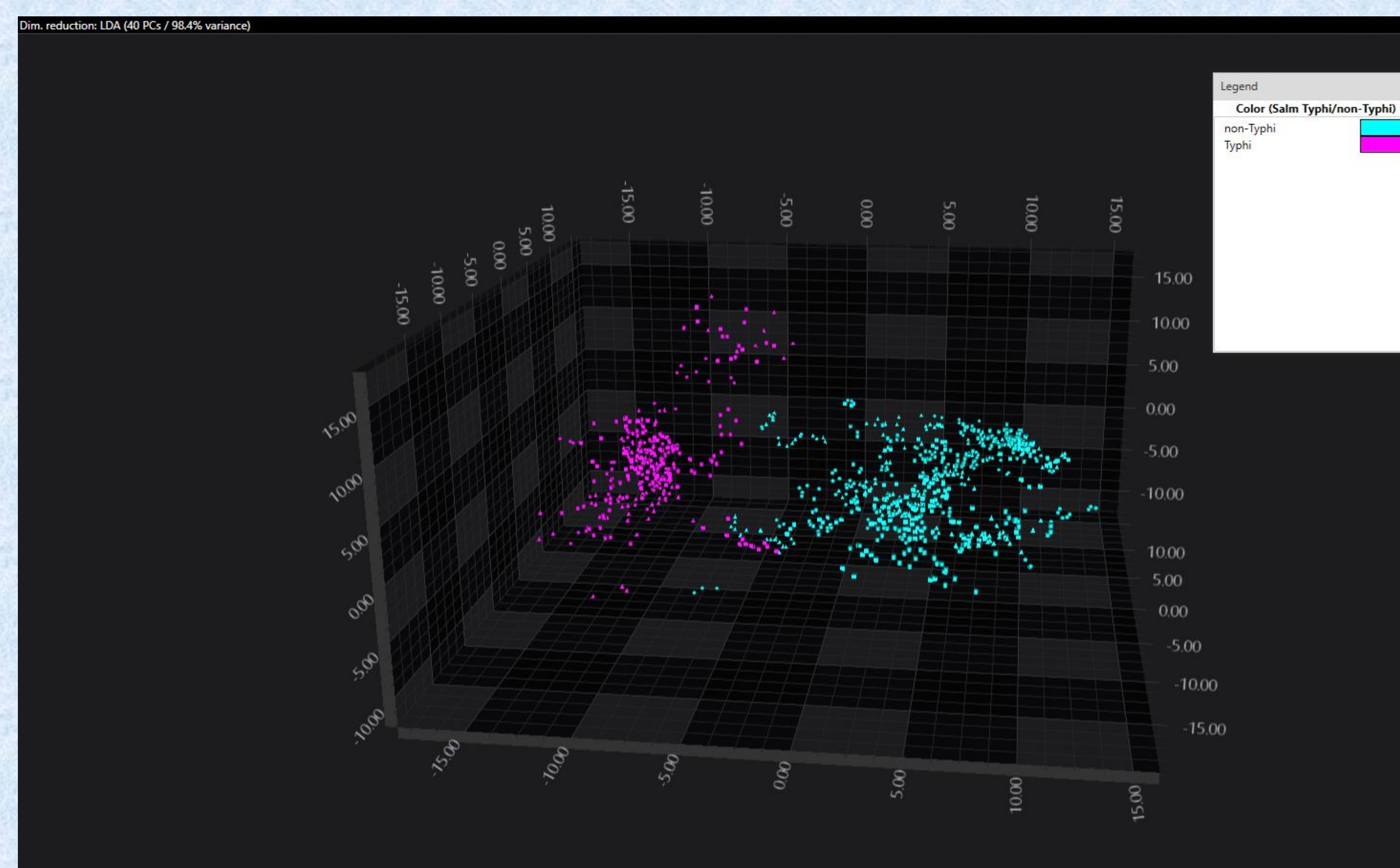


Figure 1. 3D LDA scatterplot showing the differentiation of S. Typhi (in magenta) from all the other O:9 serovars (in cyan) in the multidimensional spectral space. LDA was performed using 40 PCs, representing the 98.4% of the spectral variance. Each geometrical form represents a single spectrum.

➤ The automated classifier showed **100% accuracy**:

- **4-fold cross-validation** set (141/141)
- **Internal validation** set (52/52)
- **External independent testing** set (162/162)

3 Result Overview				
3.1 Classification Model 'Typhi-typer - Final'				
Algorithm		Training set		
Support Vector Machine (RBF)		1000		
Algorithm Version		Accuracy		
0		99.9%		
Creation date		Checksum		
2026-01-20 13:43		7229c015		
Modification date				
2026-02-03 16:48				
Isolate ID	Rep.	Best result (hits)	2nd best res. (hits)	Iso. Score
Salm D-3_Salamae-O9	9	non-Typhi (6)		100,0
Salm D-4_Enteritidis	9	non-Typhi (5)		100,0
Salm HD 1428kr24-Typhi	12	Typhi (8)		100,0
Salm QA 01BC180001451-Typhi	8	Typhi (5)		100,0
Salm S-110_Napoli	12	non-Typhi (5)		100,0
Salm S-168_Panama	3	non-Typhi (3)		100,0
Salm SZ-38_Javiana	9	non-Typhi (5)		100,0
Salm Typhi-10	3	Typhi (3)		100,0

● High reliability The score of the isolate classification is 80 or higher.
● Moderate reliability The score of the isolate classification is 20 or higher, but lower than 80.
● No reliability The score of the isolate classification is lower than 20.

Please refer to the IR Biotyper manual for details about isolate score calculation.

Figure 1. Example of the report showing the results obtained with the classifier, including the isolate ID, the number of technical replicates, the best hit result, the second-best hit (if present) and the isolate score (the summarized results of the classification of the single spectra), depicted with the corresponding color-coding.

CONCLUSION

FTIR proved to be a **robust and reliable method for the identification of S. Typhi**, representing an alternative to conventional approaches. The IR Biotyper® automated classifier is a fast and user-friendly tool, which allows a smooth implementation of FTIR results into routine workflows.