P2135

Evaluation of FT-IR spectroscopy and machine learning for the discrimination of *Escherichia coli* and *Shigella* spp.

Miriam Cordovana[1], Norman Mauder[1], Arthur B. Pranada[2], Frederik Pankok[3], Ulrike Loderstaedt[3], Simone Scheithauer[3], Denise Dekker[4], Andreas Erich Zautner[5,6], Walter Geißdörfer[7], Judith Overhoff[8], Miriam Werner[8], Andreas Wille[8], Hagen Frickmann[9,10]

[1] Bruker Daltonics GmbH & Co. KG, Bremen, Germany; [2] MVZ Dr. Eberhard & Partner Dortmund, Department of Medical Microbiology, Dortmund, Germany; [3] Department of Infection Control and Infectious Diseases, University Medical Center Göttingen, Georg August University Göttingen, Göttingen, Germany; [4] The One Health Bacteriology Group, Bernhard Nocht Institute for Tropical Medicine Hamburg, Hamburg, Germany; [5] Institute of Medical Microbiology and Hospital Hygiene, Medical Faculty, Otto-von-Guericke-University Magdeburg, Magdeburg, Germany; [6] CHaMP, Center for Health and Medical Prevention, Otto-von-Guericke-University Magdeburg, Magdeburg, Germany; [7] Institute of Microbiology – Clinical Microbiology, Immunology and Hygiene, Universitätsklinikum Erlangen, Erlangen, Germany; [8] Institute for Hygiene and Environment, City of Hamburg, Hamburg, Germany; [9] Department of Microbiology and Hospital Hygiene, Bundeswehr Hospital Hamburg, Hamburg, Germany; [10] Institute for Medical Microbiology, Virology and Hygiene, University Medicine Rostock, Rostock, Germany

Background

Certain pathovars of *Escherichia coli* (Ec) are globally leading causes of diarrhoea, foodborne outbreaks, and various extra-intestinal infections. *Shigella* spp. are enteroinvasive pathogens, causing severe gastroenteritis or even dysentery.

Indistinguishable by traditional identification methods (i.e., MALDI-TOF MS, 16S rRNA sequencing), Ec and *Shigella* spp. require specific and cumbersome biochemical tests, agglutination or PCR-based methods for their identification. Similarly, the discrimination of Ec at serotype level (lipopolysaccharide and flagellar antigens) to delineate pathogenic lineages, require serological or genomic approaches, which present some disadvantages in terms of costs, ease-of-use and applicability in routine settings.

In this study, we evaluated the discriminative power of Fourier-Transform infrared (FT-IR) spectroscopy **to distinguish *E. coli* isolates at serotype level** and to **delineate *E. coli* from *Shigella* spp**..

Material and methods

- ✓ A total of **225** well characterized strains were investigated (**n=132 *E. coli*, n=48 *S. sonnei*, n=28 *S. flexneri*, n=11 *S. boydii*, and n=6 *S. dysenteriae***). The isolates were serologically or genomically typed at species (*Shigella* spp.) and serotype (Ec) level. Among Ec, **n=71 different serotypes** were included, with different pathovars (among them, **n=22 EPEC** and **n=32 EHEC** strains)
- ✓ FT-IR spectroscopy analysis was performed by the IR Biotyper® system (IRBT - Bruker Daltonics, Germany), following the manufacturers instruction (**Figure 1)**. Three independent biological replicates on Columbia sheep blood agar incubated overnight at 35±2 ° C were included.

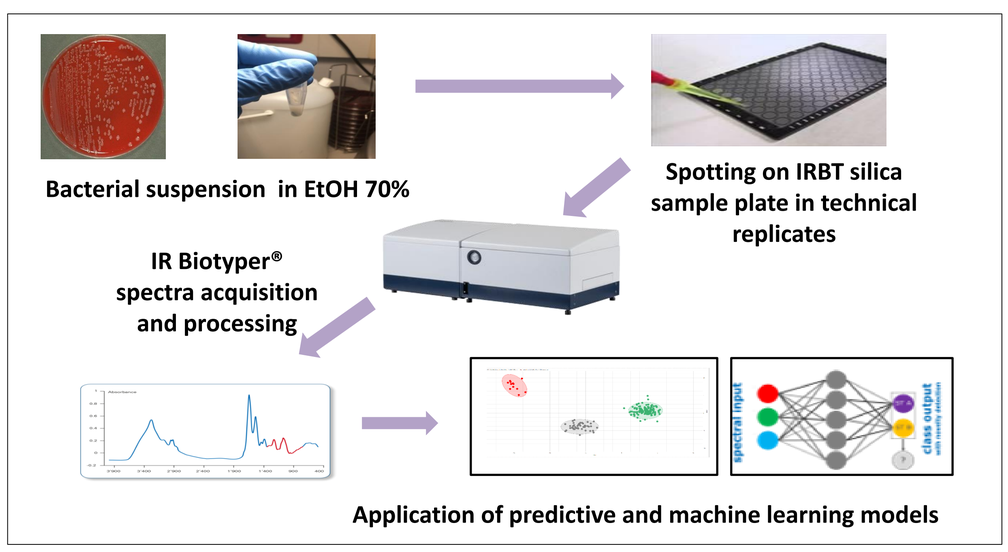

Figure 1. IR Biotyper® workflow

- ✓ **Spectra acquisition**, **processing** and **data analysis** were performed by the IR Biotyper® software V4.0.
- ✓ **Exploratory data analysis** was performed by **PCA** (Principal components analysis) and **LDA** (linear discriminant analysis).
- ✓ **LDA and different machine learning algorithms** were applied to create **predictive and learning models.** The training set included n=123 isolates (representing all the 75 groups – Ec serotypes and *Shigella* species). The remaining n=102 isolates were used as testing set.

Results

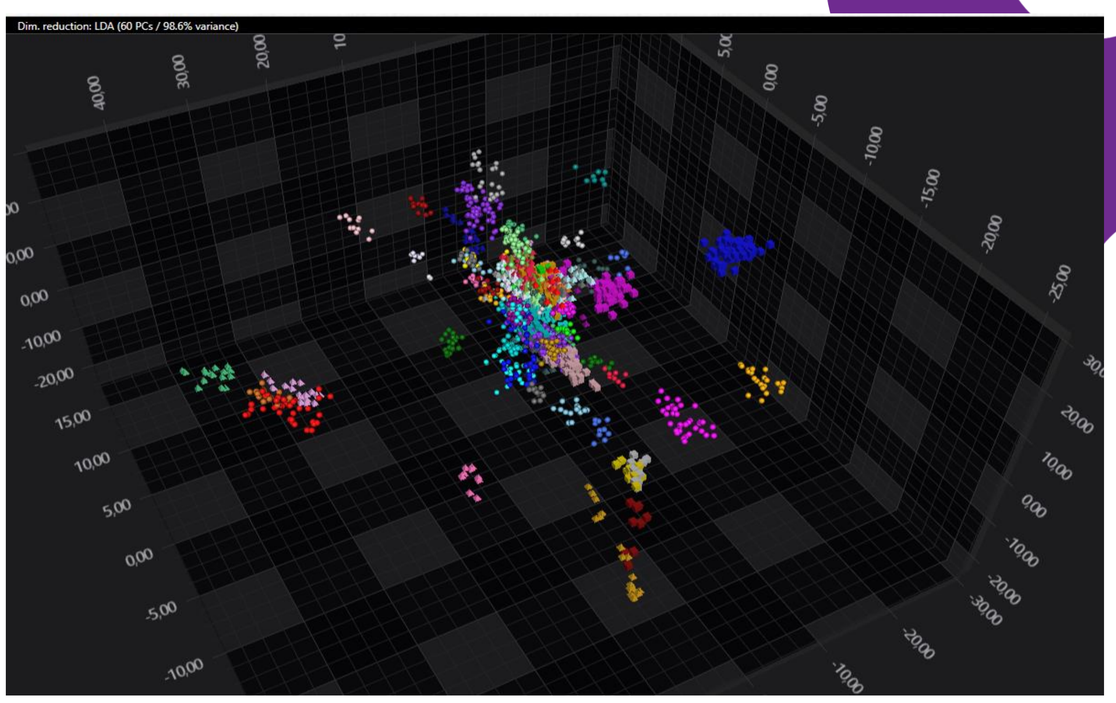

Figure 2. 3D scatter plot showing the clustering of the Ec serotypes and *Shigella* species in the multidimensional space. Each class is depicted with a different color. In the first 3 dimensions, the EHEC groups Ec O157:H7, O26:H11, the different O:15 EPEC serotypes, O45:H16, and several other non-pathogenic serovars are clearly separated.

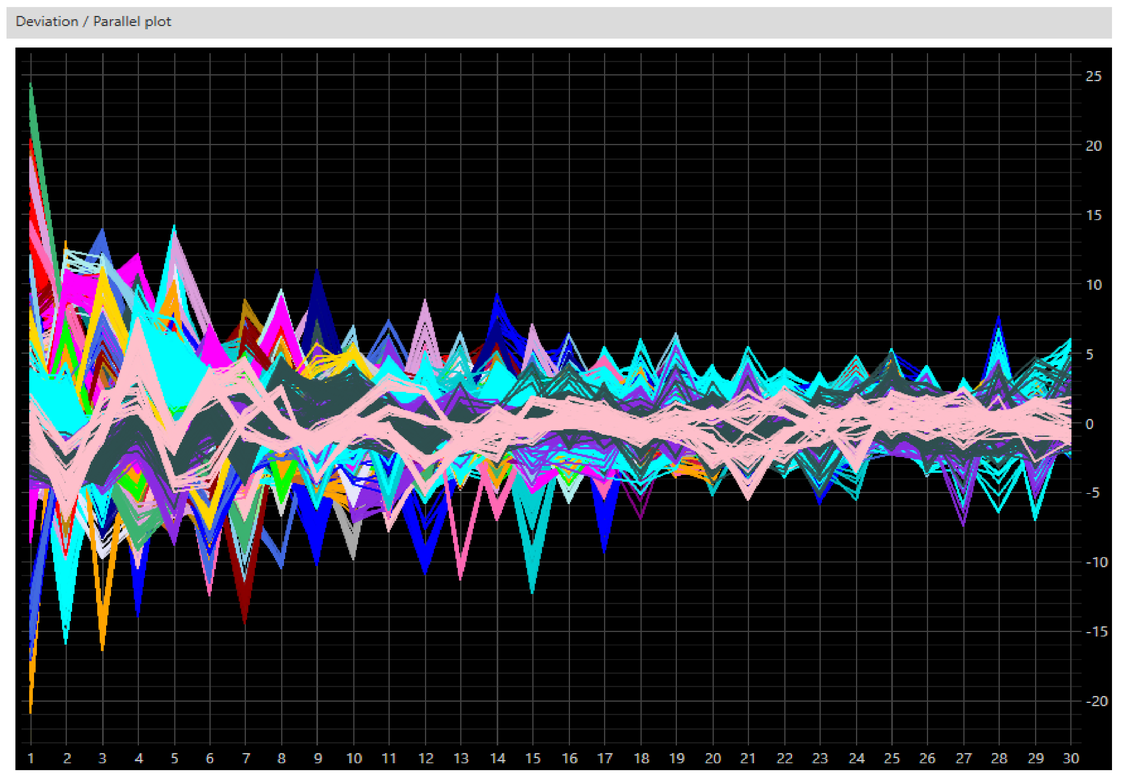

Figure 3. Deviation plot shows the separability of the samples in the first 30 dimensions, one or more of which allow the separation of all classes.

- ✓ Hierarchical cluster analysis showed that **IRBT clustering is correlated with the *E. coli* O-H serotypes and the *Shigella* species** (Adjusted Rand Index = 0.89, Adjuster Wallace Index = 0.935).
- ✓ PCA/LDA plots showed that the ***E. coli* serotypes and the 4 *Shigella* species are separable (Figures 2 and 3).**
- ✓ Preliminary results using machine learning led to an **accuracy > 99%,** nevertheless further studies, including more samples, are necessary to assess the robustness of the predictive models.

Conclusion

IR Biotyping showed the potential to delineate *E. coli* at serotype level, and to **discriminate *E. coli* from *Shigella* spp.**, demonstrating its **potential suitability for infection control, public health and epidemiological studies**. Further investigation including more strains are necessary to confirm and strengthen these promising preliminary results.

For Research Use Only - Not for clinical diagnostic usage

Contact: Miriam Cordovana (Miriam.Cordovana@bruker.com)

34th ECCMID EUROPEAN CONGRESS OF CLINICAL MICROBIOLOGY AND INFECTIOUS DISEASES
Barcelona, Spain
27 – 30 April 2024