International Journal of Food Microbiology 285 (2018) 34–41

Contents lists available at ScienceDirect

International Journal of Food Microbiology

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

Discrimination of non-typhoid *Salmonella* serogroups and serotypes by Fourier Transform Infrared Spectroscopy: A comprehensive analysis

Joana Campos[a,1], Clara Sousa[b,1], Joana Mourão[a], João Lopes[c], Patrícia Antunes[a,d], Luísa Peixe[a,*]

[a] UCIBIO/REQUIMTE, Laboratório de Microbiologia, Faculdade de Farmácia, Universidade do Porto, Portugal
[b] LAQV/REQUIMTE, Departamento de Ciências Químicas, Faculdade de Farmácia, Universidade do Porto, Portugal
[c] Research Institute for Medicines (iMed.ULisboa), Faculdade de Farmácia, Universidade de Lisboa, Lisboa, Portugal
[d] Faculdade de Ciências da Nutrição e Alimentação, Universidade do Porto, Portugal

ARTICLE INFO

Keywords:
Salmonella enterica
Typing
Fourier Transform Infrared Spectroscopy-Attenuated Total Reflectance
Multivariate data analysis
Bacterial identification methods
O-antigen

ABSTRACT

Simpler, quick and low-cost methods for routine *Salmonella enterica* typing are required for epidemiologic surveillance of this important zoonotic pathogen. In this study, using a comprehensive isolate collection, we investigated the potential of Fourier transform infrared spectroscopy (FTIRS) to discriminate the most clinically-relevant serogroups and serotypes of non-typhoid *Salmonella*. Moreover, the role of O-units composition on the FTIRS *Salmonella* discrimination was also explored.

S. enterica isolates (n = 325; 2002–2015; different sources and countries), of 57 serotypes and 15 serogroups [including the most frequent ones, B-n = 122; C-n = 108; D-n = 43 and E-n = 33)] were analysed by FTIRS. Infrared spectra were analysed by Partial Least Square Discriminant Analysis (PLSDA) and/or Principal Component Analysis (PCA).

The polysaccharides region provided the spectral sharpest differences being used in the subsequent *Salmonella* typing. Serogroups (B, C, D and E) discrimination was achieved with high accuracy (99.6% of correct assignments; PLSDA model). Differences in the O-unit structures composition of those serogroups are likely justifying the discrimination achieved. Other serogroups (G, H, K, L, M, N, O, T, U, Y, Z) were correctly predicted as not belonging to serogroups B, C, D nor E, except for 3 isolates of serogroups H (*S.* Sundsvall, n = 1) and K (*S.* Cerro, n = 2). In fact, O-unit structure of serogroup H and K shows some similarity with sub-serogroup C1 with the remaining serogroups presenting marked differences in this cellular component. The sub-serogroups discrimination was successfully achieved for C1, C2 and C3 (using PCA), and for E1-E2-E3 and E4 (by PLSDA). Appropriate serotype discrimination was obtained for most of *S.* Rissen from the remaining C1 serotypes (91.5%-PLSDA), and *S.* Enteritidis (D1) from the remaining D1/D2 serotypes (93.4%-PLSDA). The lack of available O-unit composition for particular serotypes prevents the elucidation of the role of this cellular component on the discrimination at serotype level obtained.

FTIRS was able to discriminate relevant serogroups (B, C, D and E), sub-serogroups (C1, C2 and C3; E1-E2-E3 and E4) and particular important serotypes (*S.* Enteritidis, *S.* Rissen and *S.* Senftenberg). Further studies on O-antigen composition would clarify the fundaments of discrimination obtained by FTIRS.

1. Introduction

Salmonella enterica is one of the leading causes of foodborne diseases worldwide, representing a major public health burden, even for industrialized countries of the European Union (EU) and North America (CDC, 2017; EFSA and ECDC, 2017). *S. enterica*, which comprises a quite diverse population, is transmitted to humans by a wide range of foodstuffs, mainly of animal origin, and is frequently involved in endemic or epidemic scenarios worldwide (Antunes et al., 2016; Antunes et al., 2017; Mourão et al., 2014, 2016; EFSA and ECDC, 2017). Differentiation of *S. enterica* at infra-species level is crucial, e.g. for the epidemiological investigation and control of foodborne outbreaks. Serotyping, phage typing, Multilocus Sequence Typing (MLST), Pulsed-Field Gel Electrophoresis (PFGE), Multiple Locus Variable-number

* Corresponding Author at: Laboratório de Microbiologia, Faculdade de Farmácia, Universidade do Porto, Rua de Jorge Viterbo Ferreira No 228, 4050-313 Porto, Portugal.
E-mail address: lpeixe@ff.up.pt (L. Peixe).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.ijfoodmicro.2018.07.005
Received 19 March 2018; Received in revised form 30 May 2018; Accepted 3 July 2018
Available online 10 July 2018

Tandem Repeat Analysis (MLVA), and more recently Whole-Genome Sequence (WGS) are methods widely used in *S. enterica* typing. These methods provide infra-species discrimination at different levels, although time-consuming, laborious, expensive, and may require technical expertise (Abatcha et al., 2014; Sabat et al., 2013). Remarkably, serotyping is still the globally recognized and used approach to firstly discriminate *S. enterica* isolates. This classical methodology is based on agglutination reactions with specific antisera and surface bacterial antigens, the somatic O-antigen and flagellar H-antigens (phases 1 and 2) (Grimont and Weill, 2007). The serogroup determination depends on the polysaccharide composition of O-antigen (outermost component of the cell surface lipopolysaccharide) (Grimont and Weill, 2007), which diversity is associated with the monosaccharides composition (3- to 5-sugars) and the structure of O-units (covalent bond nature and monosaccharides linkage), besides the polymerization linkage between the repeated units (2 to 8 O-units repetitions) (Davis and Mauer, 2010; Knirel, 2011; Liu et al., 2014). Additionally, the serotypes are determined based on specific combinations of O-antigen and H-antigens, being more than 2600 serotypes described (Issenhuth-Jeanjean et al., 2014). Nevertheless, few serotypes (belonging to serogroups D, B, C and E) are causing infections worldwide, namely the common *S*. Enteritidis, *S*. Typhimurium and its monophasic variant (*S.* 1,4,[5],12:i:-), and the emerging *S.* Stanley, *S*. Infantis, *S.* Rissen, *S.* Newport and *S*. Kentucky (CDC, 2017; ; EFSA and ECDC, 2017; Grimont and Weill, 2007). Despite the relevance of serotype identification, it is usually lengthy to obtain, relying on high cost and universally unavailable sera, being therefore restricted to reference laboratories. Additionally, incorrect serotypes assignments by classical approach as well as by WGS were already recognized (Abatcha et al., 2014; Sabat et al., 2013; Yachison et al., 2017).

Fourier transform infrared spectroscopy (FTIRS) is a simpler, quick and low-cost technique, which allows the discrimination of intact cells in a non-destructive manner through infrared spectra reflecting the vibrational modes of the microbial cell molecules (e.g. nucleic acids, proteins, carbohydrates, membrane and cell wall components) (Davis and Mauer, 2010). Moreover, this quick and low-cost spectroscopic technique gained more interest with the recent available system, IR Biotyper, which allows bacterial typing at routine bacteriology laboratories (www.bruker.com/applications/microbiology/strain-typing-with-ir-biotyper/features). In fact, FTIRS has been successfully explored as an alternative to discriminate diverse pathogenic bacteria causing foodborne or nosocomial infections at different infra-species level (e.g. serogroups, serotypes, clones) (Johler et al., 2016; Nyarko et al., 2014; Quintelas et al., 2018; Rebuffo-Scheer et al., 2007; Sousa et al., 2013, 2014a, 2014b; Vaz et al., 2013). Nonetheless, a comprehensive and robust assessment of the suitability of FTIRS for *S. enterica* typing by whole cell approach is lacking.

In this study, we investigated in a comprehensive and robust *Salmonella* collection, the potential of FTIRS coupled with multivariate data analysis, using whole cell analysis, to discriminate the most frequent and clinically relevant serogroups and serotypes, correlating the discrimination obtained with the O-unit composition of somatic antigen.

2. Material and methods

2.1. Bacterial collection

A total of 325 *Salmonella enterica* isolates, most belonging to *S. enterica* subsp. *enterica* (subsp. I) (n = 316) were analysed by FTIRS with Attenuated Total Reflectance (FTIRS-ATR). This collection comprises isolates from 15 serogroups, including major [B-Group O:4), n = 122; C (C1-Group O:6,7; C2-Group O:6,8; C3-Group O:8); n = 108, D (D1-Group O:9; D2-Group O:9,46), n = 43; E (E1-E2-E3-Group O:3,10, E4-Group O:1,3,19), n = 33] and less frequent ones (G-Group O:13, H-Group O:6,14, K-Group O:18, L-Group O:21, M-Group O:18, N-Group O:30, O-Group O:35, T-Group O:42, U-Group O:43, Y-Group O:48, Z-Group O:50), n = 19 (Supplementary Table S1). These serogroups encompass 57 serotypes, including the three most frequent worldwide (*S.* Enteritidis, *S.* Typhimurium and its monophasic variant, *S.* 1,4,[5],12:i:-) and others that are emergent and have clinical relevance (e.g. *S.* Rissen, *S*. Heidelberg, *S.* Infantis, *S*. Newport, *S*. Mbandaka, *S*. Stanley and *S*. Senftenberg). Forty-five less frequently detected serotypes (including ones of subspecies II-*salamae*, IIIa-*arizonae* and IIIb-*diarizonae*) in surveillance studies were also included (Supplementary Table S1). The isolates were collected between 2002 and 2015 in several countries (Portugal, Angola and Austria), being the serotypes determined by classical serotyping (Grimont and Weill, 2007) in the National Reference Centre (INSA, Lisbon, Portugal) or Austrian Agency for Health and Food Safety (AGES, Vienna, Austria). *S.* Typhimurium, *S.* 1,4,[5],12:i:- and *S.* Enteritidis were also confirmed by a PCR assay (Tennant et al., 2010). The isolates were selected from a well-characterized collection, based on clonal lineages (MLST and *Xba*I-PFGE), phenotypic/genotypic antibiotic resistance profiles and sources (human clinical cases, food products, food-animal production settings and the environment).

2.2. Spectral acquisition

The infrared spectra of studied isolates were acquired from an overnight subculture on Mueller-Hinton agar after 18 h of grow at 37 °C. The colonies were directly transferred from the agar plates to the ATR crystal (without cell component extractions), followed by air-drying on the optical surface until obtain a thin film. Spectra were acquired using a Perkin Elmer Spectrum BX FTIR System spectrophotometer equipped with a PIKE Technologies Gladi ATR accessory from 4000 to 600 cm^{-1} with a resolution of 4 cm^{-1} and 32 scan co-additions. From each isolate, nine consistent spectra were acquired, corresponding to three biological replicates (obtained in three independent days and fresh cultures) and three instrumental replicates (obtained on the same day from the same agar plate). Between each isolate measurement, a background was acquired. Due to the large amount of data generated, a mean spectrum (obtained from the nine consistent replicates) was considered in all chemometric analysis.

2.3. Spectra pre-processing and modelling

The infrared spectra were firstly processed with standard normal variate (SNV) (Naes et al., 2002), followed by the application of a Savitzky-Golay filter (15 smoothing points, 2nd order polynomial and 2nd derivate) (Savitzky and Golay, 1964). After pre-processing, infrared data were analysed by chemometric methods: non-supervised Principal Component Analysis (PCA) and the supervised Partial Least Square Discriminant Analysis (PLSDA) (Geladi and Kowalsky, 1986; Naes et al., 2002). Before modelling with PCA or PLSDA the spectra dataset was mean-centred. PLSDA modelling was preferred to PCA. PCA was used in situations where the number of isolates belonging to each class was insufficient (less than 10) to develop a robust supervised method. All multivariate chemometric data analysis were performed with Matlab software version R2016a (MathWorks, Natick, MA) and the PLS Toolbox version 8.2.1 for Matlab (Eigenvector Research, Manson, WA).

2.4. Isolates discrimination work-flow

The rational work-flow of this study consisted firstly in the development of a PLSDA model to discriminate *S. enterica* belonging to the most frequent serogroups (B, C, D and E). The developed model was further tested with additional 29 isolates ("test" isolates) belonging to 15 serogroups, modelled (B, C, D and E; n = 10) and non-modelled (G, H, K, L, M, N, O, T, U, Y, Z; n = 19) serogroups and 23 serotypes. The "test" isolates were chosen to provide a maximal diversity (Supplementary Table S1). Subsequently, the discrimination within