

Analytical Methods

Accepted Manuscript



This is an *Accepted Manuscript*, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this *Accepted Manuscript* with the edited and formatted *Advance Article* as soon as it is available.

You can find more information about *Accepted Manuscripts* in the [Information for Authors](#).

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard [Terms & Conditions](#) and the [Ethical guidelines](#) still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this *Accepted Manuscript* or any consequences arising from the use of any information it contains.

1
2
3 **Rapid discrimination of *Enterococcus faecium* strains using phenotypic**
4 **analytical techniques**
5
6

7
8 *Najla AlMasoud,^a Yun Xu,^a David I. Ellis,^a Paul Rooney,^b Jane F. Turton,^c Royston*
9 *Goodacre^{a#}*
10

11
12 ^a *School of Chemistry and Manchester Institute of Biotechnology, University of*
13 *Manchester, 131 Princess Street, Manchester, M1 7DN, UK.*
14

15
16 ^b *Belfast City Hospital, Belfast, 51 Lisburn Rd, BT9 7AB, UK.*
17

18
19 ^c *Antimicrobial Resistance and Healthcare Associated Infection Reference Unit,*
20 *National Infection Service, Public Health England, London, NW9 5EQ, UK.*
21

22
23 [#]*Correspondence to Roy Goodacre: roy.goodacre@manchester.ac.uk*
24
25
26
27

28 **Keywords:**
29

30 *Enterococcus faecium*, classification, MALDI-TOF-MS, FT-IR, Raman, pulsed-field
31 gel electrophoresis, chemometrics.
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Abstract

Clinical isolates of glycopeptide resistant enterococci (GRE) were used to compare three rapid phenotyping and analytical techniques. Fourier transform infrared (FT-IR) spectroscopy, Raman spectroscopy and matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF-MS) were used to classify 35 isolates of *Enterococcus faecium* representing 12 distinct pulsed-field gel electrophoresis (PFGE) types. The results show that the three analytical techniques provide clear discrimination among enterococci at both the strain and isolate levels. FT-IR and Raman spectroscopic data produced very similar bacterial discrimination, reflected in the Procrustes distance between the datasets (0.2125-0.2411, $p < 0.001$); however, FT-IR data provided superior prediction accuracy to Raman data with correct classification rates (CCR) of 89% and 69% at the strain level, respectively. MALDI-TOF-MS produced slightly different classification of these enterococci strains also with high CCR (78%). Classification data from the three analytical techniques were consistent with PFGE data especially in the case of isolates identified as unique by PFGE. This study presents phenotypic techniques as a complementary approach to current methods with a potential for high-throughput point-of-care screening enabling rapid and reproducible classification of clinically relevant enterococci.

1. Introduction

Enterococcus is a highly significant genus of bacteria, which causes important clinical infections including urinary tract infections (UTIs), endocarditis, meningitis, catheter-related infections, bacteremia, wound infections, pelvic and intra-abdominal infections amongst others. Some of these Gram-positive cocci were originally classified as *Streptococcus* spp. until genomic analysis by Schleifer and Kilpper-Balz in 1984 demonstrated the requirement for a separate genus classification.¹ This well-known genus is part of the normal intestinal microflora of humans and other animals.² *Enterococcus* are also part of the lactic acid bacteria (LAB) group present in foods, and whilst they are able to spoil fresh meats³, they are important in ripening and development of certain foods (i.e. dairy products), as well as being used as probiotics in humans.⁴

The majority of human clinical isolates of enterococci belong to two species, *Enterococcus faecalis* and *Enterococcus faecium*.⁵ In addition to their prevalence and pathogenicity, another very important factor associated with enterococcus is the high level of antimicrobial resistance, particularly resistance to glycopeptide antibiotics (such as vancomycin, teicoplanin and telavancin); resistant strains are referred to as GRE (glycopeptide-resistant enterococci).^{6,7}

There is a constant requirement to develop analytical methods for the discrimination of bacteria, which can be used in clinical diagnostics and food quality control. These methods should ideally be rapid, reproducible, easy to use and automated, in addition to having high resolution and sensitivity.⁸ Over a decade ago, it was common to use methods, such as polymerase chain reaction (PCR) for identification of specific DNA sequences and recognition by antibodies via enzyme-linked immunosorbent assay (ELISA), to characterize bacteria. Although these techniques are sensitive and specific and carried out using relatively inexpensive equipment, their use is limited by the complexity of preparation procedures and the requirement for specific primers and antibodies.⁹⁻¹² Nowadays, modern analytical techniques, such as matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF-MS)¹³⁻¹⁶, Fourier transform infrared (FT-IR) spectroscopy¹⁷⁻²¹ and Raman spectroscopy²²⁻²⁴ are also used for the characterization of bacteria. High dimensional and information rich datasets are produced from these techniques, which has also directly led to the requirement of robust and reliable chemometric

1
2
3 methods to assist with data deconvolution and in-depth analysis.²⁵ This saw the
4 introduction, acceptance and use of chemometrics, such as discriminant function
5 analysis (DFA)²² and hierarchical cluster analyses (HCA).²⁶⁻²⁸

6
7
8 Previously, MALDI-TOF-MS has shown promising results for bacterial
9 characterization.¹³ FT-IR and Raman spectroscopy complement each other for
10 bacterial classification; both are robust metabolic fingerprinting techniques and need
11 little sample preparation.^{29, 30} FT-IR spectroscopy is used by many researchers since
12 it is not only rapid but also offers a high-throughput and non-destructive method,
13 allowing the analysis of intact bacteria and producing unique, reproducible and
14 distinct biochemical fingerprints.³¹ Raman spectroscopy shares similar advantages to
15 FT-IR spectroscopy and also has the additional advantage of water being a very
16 weak Raman scatter³² so that samples do not need to be dried.

17
18
19 Here, the aim was to use these three distinct phenotypic approaches (namely
20 MALDI-TOF-MS, FT-IR and Raman spectroscopies) in combination with rigorous
21 chemometric analysis of the resultant datasets to classify 35 clinically relevant
22 isolates of enterococci, which had been previously typed by pulsed-field gel
23 electrophoresis (PFGE). This was carried out in order to compare the results from,
24 and determine the efficiency of, these analytical techniques for the rapid
25 differentiation of *E. faecium* strains. In future, this may allow clinical diagnostic
26 laboratories to analyze multiple bacterial samples rapidly for infection control
27 purposes in point-of-care setting within hospitals, clinics, or GP surgeries which
28 would significantly accelerate diagnosis, and potentially ensure that the correct
29 antimicrobial therapies were used if required.

30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60

2. Experimental

2.1 General chemicals

Trifluoroacetic acid (TFA), HPLC grade water, acetonitrile, sinapinic acid (SA), α -
cyano-4-hydroxycinnamic acid (CHCA), and ferulic acid (FA) were purchased from
Sigma-Aldrich (Dorset, UK).

2.2 Media

Two different types of media were used to culture the enterococci: Lysogeny Broth
(LB) and Nutrient Agar (NA). LB was prepared by mixing 5 g of yeast extract
(Amersham Life Sciences, Cleveland, USA), 10 g of tryptone (Formedia,

1
2
3 Hunstanton, UK) and 10 g of NaCl dissolved in 1 L of distilled water and the broth
4 was then autoclaved (at 121°C and 15 psi for 45 min). NA was prepared from a
5 preparatory mixture (beef extract 3 g/L, peptone 5 g/L, NaCl 8 g/L and agar 2 at 12
6 g/L) (Lab-M, Bury, UK) following the manufacturer's instructions (28 g in 1 L of
7 deionised water) and the broth was autoclaved (at 121°C and 15 psi for 15 min).
8
9

10 11 **2.3 Enterococci strains**

12 Isolates were from faecal samples from patients in a surgical ward in a hospital in
13 Belfast, UK and were collected following an increase in enterococcal infections on
14 the ward. Faecal material samples were screened onto Brilliance VRE (vancomycin-
15 resistant enterococci) agar (Oxoid, Basingstoke, UK). This agar contains antibiotics,
16 which eradicate all Gram-negative bacteria, and a high concentration of vancomycin,
17 and therefore, only vancomycin-resistant Gram-positive bacteria can grow, which
18 leads to selection of vancomycin-resistant enterococci strains. They were identified
19 as *E. faecium* by a VITEK® system (bioMérieux) and their identity confirmed by
20 MALDI-TOF analysis using a Bruker microflex instrument. The 35 isolates were
21 typed using pulsed-field gel electrophoresis (PFGE) of *Sma*I-digested genomic DNA
22 by Public Health England's National Reference Laboratory as described
23 previously.³³ Table S1 summarizes information on the 35 clinical isolates, which
24 were classified into 12 groups (12 PFGE-defined types) named: EC04, EC09, EC10,
25 EC13, EC14, EC15, EC19, EC20, UNI 156, UNI 178, UNI 191 and UNI 214, where
26 'UNI' types describe isolates that were unique within the set. All samples were
27 collected with ethical approval from the Northern Ireland Research Ethics
28 Committee, reference number "10/NIR01/20". This work did not involve any
29 experimentation on human subjects.
30
31
32
33
34
35
36
37
38
39
40
41
42

43 **2.4 Bacterial isolates**

44 The samples analyzed by the three techniques (*viz.* MALDI-TOF-MS, FT-IR and
45 Raman) were collected from the same flask to avoid any variations between different
46 preparations that may affect results obtained using the different analytical platforms.
47 First, enterococci were cultured on nutrient agar (NA) plates for 24 h at 37°C. A
48 single colony from the agar culture was used to inoculate 50 mL of Lysogeny broth
49 (LB) in a 250 mL flask which was incubated overnight at 37°C with shaking at 200
50 rpm. This was followed by measuring the optical density (OD) at 600 nm using a
51 Biomate 5 spectrophotometer (Thermo, Hemel Hempstead, UK) for each isolate.
52
53
54
55
56
57
58
59
60

1
2
3 The volume of analyzed bacterial suspension was then normalized to account for
4 variation in cell biomass in the different replicate cultures (4 biological replicates
5 were prepared for each isolate) and used to inoculate a fresh flask of broth, which
6 was incubated at 37°C for 11 h. This isolate enrichment step is required to reduce
7 interference from mixtures of different strains, which can introduce a significant
8 level of noise to readings from analytical methods. Subsequently, 10 mL from each
9 flask was collected and centrifuged at 4800 g for 10 min and the pellet washed three
10 times with sterile deionized water. Figure S1 illustrates the preparation process.

11
12 For vibrational spectroscopic analysis, the collected pellets were suspended in
13 suitable volumes of saline (0.9% (w/v) NaCl) depending on the OD (all isolates had
14 approximately the same cell density). Then, 15 µL was spotted onto a silicon plate
15 (Bruker Ltd., Coventry, UK) and was allowed to dry at 40°C for 45 min before
16 analysis with FT-IR spectroscopy. For Raman spectroscopy, 4 µL of each sample
17 was spotted onto a stainless steel plate and then allowed to dry at 40°C for 45 min.

18
19 For MALDI-TOF-MS, three different matrices were tested to find the most
20 compatible matrix with enterococci; these matrices were: FA, SA and CHCA. In
21 addition, 3 different deposition methods (sample-matrix) were tested as described
22 previously¹⁶ to find the best method for depositing the samples: mix, overlay and
23 underlay (data not shown). SA matrix and the mix deposition method were found to
24 be the optimal combination for MALDI-TOF-MS analysis for these samples. On the
25 day of analysis of the samples, the biomass was suspended in 1000 µL of 2% TFA
26 then vortexed for 3 min. An equal volume of 1 µL of bacterial suspension and matrix
27 were vortexed for 2 s and 2 µL of this mixture spotted onto a MALDI stainless steel
28 plate and allowed to dry at ambient temperature.

29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60

2.6 Fourier transform infrared (FT-IR) spectroscopy

FT-IR spectroscopy plate (Bruker Ltd., Coventry, UK) which contained 96
locations/spots was washed using 5% sodium dodecyl sulfate (SDS) solution. This
was followed by washing the plate using deionized water and allowing it to dry at
room temperature.³⁴ High-throughput screening (HTS) was carried out using a
Bruker Equinox 55 FT-IR spectrometer. The HTXTM module described by Winder *et*
*al.*³⁵ was used with this instrument. Transmission mode was used to analyze the
dried biomass to produce FT-IR spectra. The parameters used for FT-IR analysis

1
2
3 included the following: spectra were collected in the wavenumber range between
4 4000 and 600 cm^{-1} , resolution was 4 cm^{-1} and each spectrum was the average of 64
5 co-adds. Spectral acquisition and subtracting the background were achieved using
6 Opus software (Bruker Ltd.). Four biological replicates, each in four analytical
7 replicates, were analyzed and analysis was performed in three machine runs,
8 resulting in 1680 FT-IR spectra.

13 **2.7 Raman Spectroscopy**

14 This was carried out using a confocal Raman system (inVia, Renishaw plc., Wotton-
15 Under-Edge, UK) coupled with a 785 nm wavelength laser. A power intensity of ~30
16 mW was applied on the samples at an exposure time of 20 s. Four biological
17 replicates and seven different locations within each sample spot were analyzed,
18 resulting in a total of 980 Raman spectra.

23 **2.8 MALDI-TOF-MS**

24 The enterococci isolates were analyzed using an AXIMA-Confidence MALDI-TOF-
25 MS (Shimadzu Biotech, Manchester, UK), equipped with a nitrogen pulsed UV laser
26 with a wavelength of 337 nm. The parameters of this device were set as follows: 90
27 mV laser power, 91 acquired profiles with each profile containing 20 shots, linear
28 TOF, positive ionization mode, and mass-to-charge (m/z) range of 1,000-18,000. The
29 spectra were collected using a circular raster pattern. The MALDI-TOF-MS device
30 was calibrated using a protein mixture: insulin (5,735), cytochrome c (12,362), and
31 apomyoglobin (16,952) (Sigma-Aldrich). Each of 4 biological replicates from the 35
32 isolates was analyzed in four technical replicates on four different days; this led to
33 the generation of a total of 560 MALDI-TOF-MS spectra (35 isolates \times 4 biological
34 replicates \times 4 analytical replicates).

43 **2.9 Data analysis**

45 **2.9.1 Data pre-processing**

46 Opus software was used to export FT-IR data into ASCII format; the data were then
47 transferred into MATLAB 2012a (The Mathworks Inc., MA, US). All FT-IR spectra
48 were baseline corrected using standard normal variate (SNV) to remove any light
49 scattering effect. The analytical replicates were then averaged to reduce the number
50 of redundant samples. Due to the large number of samples, 8 separate (96 spot
51 silicon) sampling plates were used; therefore, it was necessary to correct for the
52 subtle differences in signals from different silicon plates. This was achieved by using
53 a piece-wise direct standardization (PDS) model.³⁶ The PDS model was built on two
54
55
56
57
58
59
60

1
2
3 different 'reference' isolates which were spotted on every plate. The pre-processed
4 FT-IR spectra were then subjected to multivariate analysis (MVA, see below).
5 Raman spectra were also normalized using standard normal variate (SNV) and then
6 subjected to MVA.
7
8

9
10 MALDI-TOF-MS data were pre-processed as follows: (i) the baseline was corrected
11 using asymmetric least squares (AsLS)³⁷, and (ii) spectra were normalized by
12 dividing each individual baseline corrected spectrum by the square root of the sum of
13 squares of the spectrum.³⁸ The pre-processed MALDI-TOF-MS data were subjected
14 to the same data analysis flow as Raman and FT-IR spectral data.
15
16
17
18

19 **2.9.2 Multivariate data analysis**

20 A flowchart of multivariate data analysis is provided in Figure 1. For all three
21 datasets, two types of classification were performed: one at the strain level (i.e. 12
22 classes) defined by PFGE, and the other at the isolate level (i.e. 35 classes, one for
23 each isolate).
24
25
26
27

28 For cluster analyses, principal components-discriminant function analysis (PC-DFA)
29 ³⁹⁻⁴¹ was applied to reduce the dimensionality of the data and discriminate samples
30 from the designated classes. The PC-DFA scores of each class were then averaged
31 and subjected to hierarchical cluster analysis (HCA).⁴² Dendrograms from each
32 analysis were generated to illustrate the relative relatedness of these bacteria.
33
34
35
36

37 Partial least squares-discriminant analysis (PLS-DA)⁴³, with 1,000 bootstrapping
38 validations⁴⁴, was also applied to obtain a validated supervised classification model
39 for discriminating different strains or isolates. In each bootstrapping process, the data
40 were randomly split into two different sets: a training set and a test set. A PLS-DA
41 model was trained on the training set and then applied to the test set to predict the
42 class membership of the samples in the test set. This process was repeated 1,000
43 times and the results were recorded and averaged to produce a $c \times c$ confusion matrix
44 (c is the number of designated classes, either 12 (strains) or 35 (isolates)), in which
45 the element at the i^{th} row, j^{th} column is the percentage of samples in class i being
46 predicted as class j on average. In order to assess the statistical significance of the
47 predictive performance of the PLS-DA models, a corresponding permutation test
48 within each bootstrapping resampling was also performed. This means that in
49 addition to building the PLS-DA model using the known class membership, another
50
51
52
53
54
55
56
57
58
59
60

1
2
3 model (called the 'null' model) was also built using a randomly permuted class
4 membership. The results of the null model were also recorded and from this the null
5 distribution was obtained. An empirical p -value was calculated by counting the
6 number of cases where the null model had obtained better predictive accuracy than
7 the real model and dividing the obtained number by the total number of
8 bootstrapping resampling (i.e. 1,000 in this study).

9
10
11 Finally, similarities between the three different datasets (FT-IR spectroscopy, Raman
12 spectroscopy and MALDI-TOF-MS data) were measured using Procrustes
13 analysis.⁴⁵ Procrustes analysis is an excellent approach for assessing the differences
14 and similarities between different ordination space from cluster analyses and has
15 been used previously for the assessment of different analytical techniques.⁴⁶ The
16 distances were calculated based on the averaged PC-DFA scores for the biological
17 replicates.
18
19
20
21
22
23
24

25 **3. Results and discussion**

26
27
28 Table S1 shows all 35 isolates belonging to 12 strains (PFGE-defined 12 types)
29 including: EC04, EC09, EC10, EC13, EC14, EC15, EC19, EC20 UNI 156, UNI 178,
30 UNI 191 and UNI 214. These strains were previously confirmed to belong to *E.*
31 *faecium* using a VITEK[®] system and by MALDI-TOF analysis using a Bruker
32 Microflex system (data not shown). The PFGE results (Fig. S2) were compared to
33 results obtained in this study using FT-IR spectroscopy^{17, 30, 46-49}, Raman
34 spectroscopy^{25, 30, 50, 51} and MALDI-TOF-MS.^{13, 14, 16, 52-54} We believe that these
35 analytical techniques in combination with chemometrics offer an improvement in the
36 classification of bacteria due to their higher biochemical resolution.
37
38
39
40
41
42
43

44 **3.1 Classification using FT-IR spectroscopy.**

45
46 In this study, four biological replicates of bacterial isolates were analyzed in four
47 analytical replicates and analysis was performed in three machine runs, resulting in a
48 total of 1680 FT-IR spectra. The three machine replicate measurements were
49 performed in order to evaluate the reproducibility of the FT-IR technique. Typical
50 spectra based on four biological replicates of representatives of 12 strains from
51 enterococcus (EC04, EC09, EC10, EC13, EC14, EC15, EC19, EC20, UNI 156, UNI
52 178, UNI 191 and UNI 214) are provided in Figure S3A. The infrared spectra
53 contain different distinct regions that can be used to characterize bacterial samples.
54
55
56
57
58
59
60

1
2
3 These have been well documented previously and include: wavenumbers around
4 3400-2850 cm^{-1} corresponding to fatty acids, at 1705-1454 cm^{-1} related to amide I
5 and II regions attributed to peptides and proteins, and around 1085-1052 cm^{-1}
6 corresponding to polysaccharides.^{19, 55, 56}
7

8
9
10 Discrimination between the strains based on visual inspection of the spectra was
11 difficult¹⁷ because these strains are very similar phenotypically. Therefore, in order
12 to develop a classification model to distinguish between bacterial samples based on
13 similarities in the spectral data, multivariate analysis was used to reduce the high
14 dimensionality of the data. First, PC-DFA was applied using 40 principal
15 components (PC) to the 12 strains (i.e. 12 classes) and 35 isolates (i.e. 35 classes)
16 using the pre-processed FT-IR spectra (Fig. 2A and 3A, respectively). Figure 2A
17 shows a clear separation between the 12 strains, displaying 4 main clusters; Cluster 1
18 is a single-member cluster (SMC) containing only (EC10), Cluster 2 includes (EC20
19 and UNI 156), Cluster 3 (UNI 191, EC04 and EC15) and Cluster 4 formed a large
20 group and is a combination of (EC13, EC19, EC14, EC09, UNI 214 and UNI 178).
21 Each cluster is represented by a different color in the figure. As described above,
22 HCA was undertaken using spectral data in order to simplify the DFA plot and to
23 illustrate the related strains. Cluster analysis was based on averaged DFA scores (12
24 classes/strains), using Ward's linkage as shown in Figure 2B. Clusters seen in Figure
25 2A are reflected in the HCA dendrogram plot (Fig. 2B).
26
27

28
29
30 PC-DFA was subsequently performed for all the 35 isolates and the results are
31 provided in Figure 3. Clear separation between all 35 isolates was observed despite
32 the fact that there were a much higher number of classes to be separated than the
33 number of strains. For example, clear separation was observed between the two
34 representatives of EC10 (139 and 151). Furthermore, results generated using PFGE
35 correlated well with FT-IR spectroscopic data. For example, the UNI 156 and
36 UNI 178 were seen as unique by both techniques. In addition, the three EC20
37 isolates (192, 198 and 204) and EC19 isolates (173, 174 and 175) clustered together
38 and were not differentiated using FT-IR spectroscopy, which was also observed in
39 the PFGE results, where the bands were quite similar (Fig. 3B). This implies that the
40 isolates within each of these groups are highly similar to each other phenotypically
41 and genetically. Finally, two more clusters were observed, with one cluster
42 containing all the EC04, EC15 and UNI 191 strains and the remainder of the isolates
43 forming another cluster.
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 The PLS-DA classification using FT-IR spectral data achieved an average correct
4 classification rate (CCR) of 89.4% at the strain level and 54.3% at the isolate level,
5 both with an empirical p -value of <0.001 , i.e. not a single case where the null model
6 obtained better results, indicating that the predictive accuracies were highly
7 significant. The null distributions are provided in Figure S4A and B at the two
8 levels.
9

10
11
12 The confusion matrices of strains and isolates classification are presented in Table 1
13 and Table S3, respectively. Most of the 12 strains showed high prediction
14 accuracies, for example EC04, EC10, EC13 and EC20 had accuracies of 89.9%,
15 99.7%, 99.8% and 99.2%, respectively. However, EC14 and UNI 214 had lower
16 prediction accuracies of 47.3% and 58.9%, respectively. The confusion matrix
17 showed that there was a certain level of overlap between (EC14 and EC09) and (UNI
18 214 and EC19).
19

20
21 Furthermore, in-depth analysis of the confusion matrix (Fig. 4) showed that
22 classification of unique strains was generally in line with PFGE results. In Figure 4,
23 high percentage class membership assignments are represented by warm colors (e.g.
24 red), indicating agreement between predicted classes and known classes. It is also
25 interesting to see that representatives from EC19 and EC20 formed two “squares” of
26 “tiles” on the diagonal line, in which the colors were similar to each other. Results
27 from Figure 4 suggest that the PLS-DA model was not able to differentiate the
28 isolates within EC19 and EC20, yet another observation that is consistent with PFGE
29 results. On the other hand, all representatives of EC04 and EC09 (160 and 133) were
30 unique in the FT-IR spectroscopy profile using the PLS-DA model but had visually
31 similar PFGE profiles. This is most likely due to PFGE providing genetic
32 information^{57, 58} while FT-IR spectroscopy describes phenotypes.^{27, 59} This implies
33 that isolates from EC19 and EC20 are highly conserved phenotypically, whereas
34 those from EC04 and EC09 are not, and such subtle differences in phenotypes were
35 detected by FT-IR spectroscopy. Our observations showed that FT-IR spectroscopy
36 appears to be a very promising analytical approach for discrimination of enterococci
37 at different levels. In line with the results presented in this study, work carried out by
38 Guibet *et al.* showed that clear discrimination and classification of enterococci
39 strains can be achieved using FT-IR spectroscopy.^{60, 61}
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55

56 **3.2 Classification using Raman spectroscopy.** In addition to the FT-IR
57 spectroscopy technique used in this study, Raman spectroscopy was used as a
58
59
60

1
2
3 complementary technique.^{17, 61-63} As expected, the two techniques generated
4 different spectra. These two approaches are complementary due to the selection
5 rules, whereby infrared causes a change in the net dipole moment in a particular
6 functional group, induced by molecular vibrations, whereas Raman causes a change
7 in the polarization of bonds within a molecule. Therefore, bonds within a molecule
8 are generally infrared or Raman active with the result being that the two techniques
9 can provide complementary (bio) chemical information.^{29, 64}
10
11
12
13
14
15

16 Raman spectra of the 12 *E. faecium* strains are shown in Figure S3B. Raman spectra
17 for these types appeared almost indistinguishable and no differences were detected
18 on visual inspection. Moreover, some specific peaks which were identified in these
19 spectra included: peaks at around 722 cm⁻¹, 783 cm⁻¹, 854 cm⁻¹, 1004 cm⁻¹, 1098 cm⁻¹,
20 1334 cm⁻¹, 1451 cm⁻¹ and 1664 cm⁻¹, which correspond to adenine, cytosine/uracil,
21 tyrosine, phenylalanine, phosphate, guanine, protein and amide I, respectively.⁶⁵⁻⁶⁷
22
23
24
25
26

27 PC-DFA scores plot of pre-processed Raman spectra for the 12 PFGE-defined types
28 is shown in Figure S5A. The figure shows classification results similar to those seen
29 with FT-IR spectroscopy data. There was an obvious overlap between the two
30 spectroscopic techniques, especially with representatives of EC10. However, EC20
31 overlapped with UNI 156 in FT-IR spectroscopy data, whereas EC20 was closer to
32 UNI 178 based on Raman spectroscopy data. These observations can be seen in the
33 HCA dendrogram based on Raman data (Fig. S5B), which was quite similar to the
34 HCA results generated from FT-IR data. Looking back at the dendrogram in
35 Figure S2 based on PFGE data, visual inspection showed that there were some
36 similarities between results generated via spectroscopic techniques and those based
37 on PFGE; for example, EC04 and EC15 were shown to overlap in both sets of results
38 (Fig. S2).
39
40
41
42
43
44
45
46
47

48 As with FT-IR data, Raman spectroscopy data on the 35 isolates were also analyzed
49 using to PC-DFA and HCA (Fig. S6A and B, respectively). The results suggested
50 that Raman spectroscopy was also successful in discriminating the two
51 representatives of EC10 (139 and 151), which was also the case using FT-IR
52 analysis (Fig. 3). Furthermore, in order to ensure the classification is robust, the data
53 were analyzed using a heat map based on PLS-DA (Fig. S6C). The results suggested
54
55
56
57
58
59
60

1
2
3 that all the isolates indicated as unique (UNI) by PFGE were also unique in the PLS-
4 DA model generated using Raman spectroscopy data.
5
6

7 In addition, chemometric-based identification was carried out using PLS-DA at both
8 the strain and isolate levels and the predictive accuracies were calculated based on
9 1,000 bootstrapping resampling using Raman spectral data. The null distribution was
10 obtained (Fig. S4C and D) at both the strain (12 classes) and isolate levels (35
11 classes) resulting in average CCR of 69.3% ($p < 0.001$) and 21.1% ($p < 0.001$),
12 respectively. The CCR from FT-IR data was higher at both levels compared to
13 Raman data possibly due to the higher reproducibility of FT-IR data. Confusion
14 matrices were also generated at both the strain level (Table S2A) and the isolate
15 level (data not shown); these results suggested that Raman spectroscopy can also be
16 used as a robust technique for bacterial discrimination. In-depth analysis showed that
17 Raman spectroscopy generated around 70% prediction accuracy at the strain level
18 which is lower than that of FT-IR spectroscopy (nearly 90%). This is most likely due
19 to the low concentration of cells used for analysis: the infrared interrogation beam
20 used was *ca.* 1 mm and passes completely through the dried bacterial film; while the
21 Raman microscope delivers a highly focussed laser beam with an interrogation
22 volume of ~ 1 pL and therefore measures very few bacteria. To overcome this
23 limitation with Raman, bacteria can be analyzed directly from the agar plates or
24 surface-enhanced Raman spectroscopy (SERS) as an alternative technique⁶⁸⁻⁷⁰, but
25 this is an area for future study.
26
27
28
29
30
31
32
33
34
35
36
37
38

39 **3.3 Classification using MALDI-TOF-MS.** As described in the Materials and
40 Methods section, four biological replicates were analyzed in four analytical
41 replicates for each bacterial strain, resulting in 560 MALDI-TOF-MS spectra; both
42 the biological and technical replicates clustered closely together ensuring good
43 bioanalytical reproducibility (data not shown). The spectra for the 35 enterococci
44 isolates were pre-processed before data analysis. The typical pre-processed positive
45 ion mode MALDI-TOF-MS spectra for all 12 *Enterococcus* strains (EC04, EC09,
46 EC10, EC13, EC14, EC15, EC19, EC20, UNI 156, UNI 178, UNI 191 and UNI 214)
47 are provided in Figure S3C. In general, the MALDI-TOF-MS spectra were of high
48 quality with high signal-to-noise ratios in the acquisition m/z range 1,000-18,000 and
49 a high number of peaks for each studied strain were detected. There are many factors
50 that can affect MALDI-TOF-MS results and some of these can differ from lab to
51
52
53
54
55
56
57
58
59
60

1
2
3 another, such as the type of medium used ⁷¹, sample handling, type of matrix ⁷²,
4 sample deposition method ⁷³, solvents, instrument settings ^{74, 75} and the type of data
5 analysis chosen. ^{41, 76} These can inadvertently affect MALDI-TOF-MS results and
6 subsequent PC-DFA and HCA.
7
8

9
10 MALDI-TOF-MS spectra are not readily interpretable from the 35 isolates as they
11 are similar phenotypically and MALDI-TOF-MS spectra show only two dimensions
12 ($m/z \times$ intensity). Therefore, as is the case for the vibrational spectroscopy
13 techniques, robust multivariate analysis methods were employed for this purpose.
14 The results of PC-DFA using 12 classes (12 strains) in a three-dimensional plot of
15 DF1 vs DF2 vs DF3 and a two-dimensional plot of DF2 vs DF3 are shown in Figure
16 S7A and B, respectively. Four main clusters were observed in the PC-DFA plots;
17 SMC (Cluster) 1 contains only UNI 178; Cluster 2 contains EC20; Cluster 3 consists
18 of EC04, EC10, EC15 and UNI 191; and Cluster 4 formed a large group of (EC13,
19 EC19, EC14, EC09, UNI 214 and UNI 156). Results from the HCA dendrogram
20 (Fig. S7C) confirmed the separation between the 12 classes (i.e. 12 strains). This
21 indicated that UNI 178 is phenotypically very different from the other strains based
22 on MALDI-TOF-MS data.
23
24

25
26 PC-DFA was also applied to data from the 35 isolates; the results showed that
27 isolates number 160 and 219 (both from EC09) were very different from the other
28 isolates. Therefore, another PC-DFA was carried out with these two outliers
29 removed and the HCA results are shown in Figure S8D. It appears that all
30 representatives of EC20 (204, 198 and 192) overlap with each other, which was also
31 observed in FT-IR and Raman spectroscopy data, with the exception that isolate 192
32 slightly differed from the other two representatives (204 and 198) in the HCA
33 dendrogram when using Raman data (Fig. S6B). However, analysis by PFGE
34 showed that isolates 192 and 198 clustered more closely with each other than with
35 isolate 204.
36
37

38
39 Furthermore, PLS-DA model applied to MALDI-TOF-MS data achieved an average
40 CCR of 78.2% ($p < 0.001$) and 35.7% ($p < 0.001$) for the 12 (strains) and 35 (isolates)
41 classes, respectively. When PLS-DA was undertaken with 33 isolates (with isolates
42 160 and 219 removed), the average CCR for the isolates increased to 53.95%
43 ($p < 0.001$). The prediction accuracies for the 12 classes are shown in Table S2B and
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 those for the 35 classes (isolates) are shown in Table S4. Table S2B shows that
4 discrimination between most of the strains (12 classes) using MALDI-TOF-MS data
5 achieved high correct classification rates, except for EC14 and UNI 191, which had
6 rather low classification rates. Confusion matrices for the 35 classes and the 33
7 classes (160 and 219 isolates removed) are shown in Figure S8A and C, respectively.
8 From these matrices, it can be seen that all the isolates identified by the reference
9 laboratory as unique (UNI), which included isolates 156, 178, 191 and 214, were
10 also classified as unique based on MALDI-TOF-MS data. Moreover, EC20 and
11 EC19 were assigned the same classification in PFGE typing, and this was in
12 agreement with MALDI-TOF-MS, FT-IR spectroscopy and Raman spectroscopy
13 data. In addition, based on MALDI-TOF-MS data (Fig.S8A and C), representatives
14 of EC13 (152, 154 and 155) belonged to the same cluster, and isolates 177 from
15 EC13 was significantly different from the remaining EC13 strains; this was also
16 observed in FT-IR and PFGE data. Looking back at Figure S8C, it can be seen that
17 all the strains from EC04 were unique in MALDI-TOF-MS and FT-IR profiles when
18 using PLS-DA modelling.
19
20
21
22
23
24
25
26
27
28
29

30 **3.4 Procrustes distance test of the three analytical techniques.** Analytical
31 techniques such as FT-IR spectroscopy, Raman spectroscopy and MALDI-TOF-MS
32 are currently used in clinical research studies worldwide and many reports have been
33 published showing advantages of using such techniques.^{24, 54, 77, 78}
34
35
36
37
38

39 Compared to PCR, the ‘gold standard’ technique for enterococci identification, no
40 conclusive evidence was identifiable in the literature. Application of PCR to the
41 classification of *E. coli* from five different sources recorded average correct
42 classification of 84% (Seurinck et al., 2003), a level comparable to that obtained
43 with FT-IR spectroscopy in this study (89.4%) albeit with a different microorganism.
44 Kirschner *et al.*⁶¹ demonstrated accurate identification and classification of 18
45 strains from 6 different species belonging to enterococci using vibrational
46 spectroscopic techniques in combination with chemometrics. This study suggested
47 that FT-IR and Raman spectroscopies can offer potential alternatives to the
48 conventional typing tests, based on PCR, due to their speed and ease of use,
49 demonstrating high consistency between classifications based on FT-IR and Raman
50 methods. Based on comparison with classification using PCR, they advocated the
51
52
53
54
55
56
57
58
59
60

1
2
3 use of FT-IR and Raman techniques due to the limited number of Enterococci
4 species that could be analysed by PCR and the requirement for very specific
5 procedures, which makes these analytical techniques more suitable for routine use.
6
7
8 Our results are in agreement with their findings; however, Kirschner *et al.*⁶¹ did not
9 report comparative analysis of correct classification rates from these techniques.
10
11 Oliveira *et al.*⁵¹ showed that Raman spectroscopy, in combination with a
12 chemometric algorithm, can be used to discriminate between seven different colonies
13 of Gram-positive and Gram-negative bacteria. In another previous study, it was also
14 shown that 59 clinical bacterial strains associated with urinary tract infections (UTIs)
15 could be identified using FT-IR and Raman spectroscopy.¹⁷ As an alternative to
16 vibrational spectroscopic techniques, MALDI-TOF-MS is a relatively new technique
17 which has shown very promising results for identification in agreement with
18 methodologies carried out in microbiological laboratories, and therefore has been
19 used for the identification and classification of bacterial species^{15, 79, 80} and is
20 appearing in many clinical microbiology testing laboratories.^{54, 81, 82}

21
22
23
24
25
26
27
28 Previous studies have generally focussed on the application of just one or two
29 analytical techniques for the classification of *Enterococcus* spp. However, to
30 generate complementary data and more comprehensive analysis, this study combines
31 three different analytical techniques – FT-IR spectroscopy, Raman spectroscopy and
32 MALDI-TOF-MS – to analyze whole bacterial cells. Successful classification was
33 demonstrated at the strain (i.e. 12 classes) and isolate (i.e. 35 classes) levels based on
34 data generated by the three analytical platforms. In order to assess the overall
35 information content in the spectra that has been revealed by the cluster analysis from
36 the scores plots, Procrustes analysis was employed to assess the overall similarity
37 between the patterns detected by these three platforms. The results are presented in
38 terms of Procrustes distance (Table 2A and B), where the Procrustes distance varies
39 from 0 to 1; the lower the distance, the higher the similarity between the results. The
40 comparisons were made using averaged PC-DFA scores. For each dataset, there
41 were two sets of PC-DFA scores, one at the strain level (12 classes) and another for
42 isolates classification (35 classes). For each set of PC-DFA scores, the scores were
43 then averaged according to their strain label and isolate label to give two sets of
44 *averaged* PC-DFA scores.

45
46
47
48
49
50
51
52
53
54
55
56 The findings in Table 2 can be summarized as follows:
57
58
59
60

- 1
2
3 (i) The patterns in the PC-DFA scores at strain and isolate levels were highly
4 similar to each other for all the three analytical platforms. The Procrustes
5 distances varied from 0.0681 to 0.1812. This suggested that the variation
6 originating from different bacteria is the main factor in PC-DFA, i.e. the
7 differences between different strains were significantly higher than those
8 between different isolates.
9
10 (ii) The two vibrational spectroscopic techniques (FT-IR and Raman) generated
11 highly similar results both at the strain and isolate classification levels, with
12 the corresponding Procrustes distances varying from 0.2112 to 0.3187.
13
14 (iii) However, the results generated by MALDI-TOF-MS were significantly
15 different from those generated by the two spectroscopic techniques, and the
16 corresponding Procrustes distances were all above 0.8. Such differences can
17 be mainly attributed to data on isolate UNI 178, which appeared to be very
18 different to other strains in the MALDI-TOF-MS dataset.
19
20
21
22
23
24
25

26 Table S5 shows a summative comparison of the 4 main clusters identified based on
27 the three analytical techniques using PC-DFA plots of the 12 *E. faecium* strains (12
28 classes). It can be seen from this table that despite the large Procrustes distances
29 between data generated by MALDI-TOF-MS and those generated by the other two
30 techniques, the main identified clusters patterns observed in all three datasets were
31 still largely consistent.
32
33
34
35
36
37

38 4. Conclusions

39
40 The results obtained from the three analytical techniques (mass spectrometry and
41 vibrational spectroscopy) demonstrated that good discrimination between *E. faecium*
42 bacteria can be achieved at both the strain and isolate levels and the detected patterns
43 from these techniques were highly similar. However, UNI 178 was detected to be
44 different in MALDI-TOF-MS data, which differed from the two vibrational
45 spectroscopy techniques employed in this study.
46
47
48
49
50

51 The results obtained using these spectroscopic phenotyping approaches were mostly
52 consistent with previous results obtained from experiments carried out using the
53 genotypic classification method, PFGE. Some of the results differed when directly
54 comparing our analytical approach with results from the molecular approach and
55
56
57
58
59
60

1
2
3 these differences may be due to comparing phenotypic differences from whole-
4 organism fingerprinting with genotypic differences using PFGE.
5
6

7 In conclusion, we have assessed multiple analytical phenotypic as complementary
8 approaches to current molecular methods. All methods provided excellent clustering
9 which was in general agreement with genotypic baseline methods, as well as
10 allowing excellent discrimination to the strain level and good resolution at the sub-
11 strain level. We believe that these three different physicochemical techniques have
12 excellent potential as high-throughput point-of-care screening tools, and for the rapid
13 and reproducible classification of clinically relevant bacteria, such as *E. faecium*.
14 However, further method development may be required to optimise these methods
15 for reliable analysis of bacterial mixtures.
16
17
18
19
20
21
22

23 **Acknowledgments**

24 NM thanks The Saudi Ministry of Higher Education and Princess Nora bint Abdul
25 Rahman University for funding. YX thanks Cancer Research UK (including
26 Experimental Cancer Medicine Centre award) and Wolfson Foundation, DIE and RG
27 thank BBSRC (BB/L014823/1) for support for Raman spectroscopy.
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

References

1. K. H. Schleifer and R. Kilpper-Bälz, *Int. J. Syst. Evol. Microbiol.*, 1984, **34**, 31-34.
2. F. H. Kayser, K. A. Bienz and J. Eckert, *Medical microbiology*, Georg Thieme Verlag, 2011.
3. J. R. Hayes, L. L. English, P. J. Carter, T. Proescholdt, K. Y. Lee, D. D. Wagner and D. G. White, *Appl. Environ. Microbiol.*, 2003, **69**, 7153-7160.
4. C. M. Franz, M. E. Stiles, K. H. Schleifer and W. H. Holzapfel, *Int. J. Food Microbiol.*, 2003, **88**, 105-122.
5. M. McCracken, A. Wong, R. Mitchell, D. Gravel, J. Conly, J. Embil, L. Johnston, A. Matlow, D. Ormiston and A. Simor, *J. Antimicrob. Chemother.*, 2013, dkt054.
6. N. Woodford, *J. Med. Microbiol.*, 1998, **47**, 849-862.
7. C. A. Arias and B. E. Murray, *Nature Rev. Microbiol.*, 2012, **10**, 266-278.
8. S. Altekruze, M. Cohen and D. Swerdlow, *Emerg. Infect. Dis.*, 1997, **3**, 285-294.
9. E. Engvall, *Medical biology*, 1977, **55**, 193.
10. R. H. Yolken, *Yale J. Biol. Med.*, 1980, **53**, 85.
11. D. Ke, F. J. Picard, F. Martineau, C. Ménard, P. H. Roy, M. Ouellette and M. G. Bergeron, *J. Clin. Microbiol.*, 1999, **37**, 3497-3503.
12. D. J. Reen, *Basic Protein and Peptide Protocols*, 1994, 461-466.
13. P. Lasch, H. Nattermann, M. Erhard, M. Stämmler, R. Grunow, N. Bannert, B. Appel and D. Naumann, *Anal. Chem.*, 2008, **80**, 2026-2034.
14. M. Quintela-Baluja, K. Böhme, I. C. Fernández-No, S. Morandi, M. E. Alnakip, S. Caamaño-Antelo, J. Barros-Velázquez and P. Calo-Mata, *Electrophoresis*, 2013, **34**, 2240-2250.
15. P. Lasch, C. Fleige, M. Stämmler, F. Layer, U. Nübel, W. Witte and G. Werner, *J. Microbiol. Methods*, 2014, **100**, 58-69.
16. N. AlMasoud, Y. Xu, N. Nicolaou and R. Goodacre, *Anal. Chim. Acta*, 2014, **840**, 49-57.
17. R. Goodacre, R. Burton, N. Kaderbhai, A. M. Woodward, D. B. Kell and P. J. Rooney, *Microbiology*, 1998, **144**, 1157-1170.
18. D. Helm, H. Labischinski, G. Schallehn and D. Naumann, *J. Gen. Microbiol.*, 1991, **137**, 69-79.
19. D. Naumann, D. Helm and H. Labischinski, *Nature*, 1991, **351**, 81-82.
20. C. X. Lu, Y. A. Liu, X. H. Sun, Y. J. Pan and Y. Zhao, *Acta Chimica Sinica*, 2011, **69**, 101-105.
21. Y. Burgula, D. Khali, S. Kim, S. S. Krishnan, M. A. Cousin, J. P. Gore, B. L. Reuhs and L. J. Mauer, *J. Rapid Meth. Aut. Mic.*, 2007, **15**, 146-175.
22. E. C. López-Díez and R. Goodacre, *Anal. Chem.*, 2004, **76**, 585-591.
23. K. Maquelin, C. Kirschner, L. P. Choo-Smith, N. van den Braak, H. P. Endtz, D. Naumann and G. J. Puppels, *J. Microbiol. Meth.*, 2002, **51**, 255-271.
24. M. Beekes, P. Lasch and D. Naumann, *Vet. Microbiol.*, 2007, **123**, 305-319.
25. D. I. Ellis, D. P. Cowcher, L. Ashton, S. O'Hagan and R. Goodacre, *Analyst*, 2013, **138**, 3871-3884.
26. C. S. Gutteridge, L. Valus and H. J. H. Macfie, in *Computer-Assisted Bacterial Systematics*, ed. M. G. J. G. Priest, Academic Press, London, 1985, pp. 369-401.
27. R. Davis and L. Mauer, *Current research, technology and education topics in applied microbiology and microbial biotechnology*, 2010, **2**, 1582-1594.
28. J. P. Dworzanski, S. V. Deshpande, R. Chen, R. E. Jabbour, A. P. Snyder, C. H. Wick and L. Li, *J. Proteome Res.*, 2006, **5**, 76-87.
29. R. Goodacre, B. S. Radovic and E. Anklam, *Appl. Spectrosc.*, 2002, **56**, 521-527.
30. D. I. Ellis and R. Goodacre, *Analyst*, 2006, **131**, 875-885.
31. A. A. Argyri, R. M. Jarvis, D. Wedge, Y. Xu, E. Z. Panagou, R. Goodacre and G.-J. E. Nychas, *Food Control*, 2013, **29**, 461-470.

- 1
2
3 32. R. Goodacre, E. M. Timmins, P. J. Rooney, J. J. Rowland and D. B. Kell, *FEMS Microbiol. Lett.*, 1996, **140**, 233-239.
- 4
5 33. N. Woodford, R. Reynolds, J. Turton, F. Scott, A. Sinclair, A. Williams and D. Livermore, *J. Antimicrob. Chemother.*, 2003, **52**, 711-714.
- 6
7 34. S. A. Patel, F. Currie, N. Thakker and R. Goodacre, *Analyst*, 2008, **133**, 1707-1713.
- 8 35. C. L. Winder, S. V. Gordon, J. Dale, R. G. Hewinson and R. Goodacre, *Microbiology*, 2006, **152**, 2757-2765.
- 9
10 36. Y. Wang, D. J. Veltkamp and B. R. Kowalski, *Anal. Chem.*, 1991, **63**, 2750-2756.
- 11 37. P. H. C. Eilers, *Anal. Chem.*, 2004, **76**, 404-411.
- 12 38. R. G. Brereton, *Chemometrics: data analysis for the laboratory and chemical plant*, John Wiley & Sons, 2003.
- 13
14 39. B. F. Manly, *Multivariate statistical methods: a primer*, CRC Press, 2004.
- 15 40. G. G. Harrigan, R. H. LaPlante, G. N. Cosma, G. Cockerell, R. Goodacre, J. F. Maddox, J. P. Luyendyk, P. E. Ganey and R. A. Roth, *Toxicology Letters*, 2004, **146**, 197-205.
- 16
17
18 41. P. S. Gromski, H. Muhamadali, D. I. Ellis, Y. Xu, E. Correa, M. L. Turner and R. Goodacre, *Anal. Chim. Acta*, 2015, **879**, 10-23.
- 19
20 42. T. Hastie, R. Tibshirani and J. Friedman, New York: Springer, 2009.
- 21 43. M. Barker and W. Rayens, *J. Chemometrics*, 2003, **17**, 166-173.
- 22 44. B. Efron and R. J. Tibshirani, *An introduction to the bootstrap*, CRC press, 1994.
- 23 45. J. C. Gower and G. B. Dijkstrahuis, *Procrustes problems*, Oxford University Press Oxford, 2004.
- 24
25 46. P. R. Peres-Neto and D. A. Jackson, *Oecologia*, 2001, **129**, 169-178.
- 26 47. H. AlRabiah, Y. Xu, N. J. Rattray, A. A. Vaughan, T. Gibreel, A. Sayqal, M. Upton, J. W. Allwood and R. Goodacre, *Analyst*, 2014, **139**, 4193-4199.
- 27
28 48. D. Naumann, *Infrared Physics*, 1984, **24**, 233-238.
- 29 49. D. Naumann, V. Fijala, H. Labischinski and P. Giesbrecht, *J. Mol. Struct.*, 1988, **174**, 165-170.
- 30
31 50. L. Mariey, J. P. Signolle, C. Amiel and J. Travert, *Vib. Spectrosc.*, 2001, **26**, 151-159.
- 32
33 51. A. Wokaun, *Berichte der Bunsengesellschaft für physikalische Chemie*, 1996, **100**, 1268-1268.
- 34
35 52. H. Muhamadali, M. Chisanga, A. Subaihi and R. Goodacre, *Anal. Chem.*, 2015, **87**, 4578-4586.
- 36
37 53. C. Benagli, V. Rossi, M. Dolina, M. Tonolla and O. Petrini, *Plos One*, 2011, **6**, e16424
- 38
39 54. R. Cramer, J. Gobom and E. Nordhoff, *Expert Rev Proteomics*, 2005, **2**, 407-20
- 40 55. D. I. Ellis, W. B. Dunn, J. L. Griffin, J. W. Allwood and R. Goodacre, *Analyst*, 2007, **131**, 875-885.
- 41
42 56. E. Carbonnelle, C. Mesquita, E. Bille, N. Day, B. Dauphin, J.-L. Beretti, A. Ferroni, L. Gutmann and X. Nassif, *Clinical Biochemistry*, 2011, **44**, 104-109.
- 43
44 57. S. Kim, B. L. Reuhs and L. J. Mauer, *J. Appl. Microbiol.*, 2005, **99**, 411-417.
- 45 58. D. I. Ellis, G. G. Harrigan and R. Goodacre, in *Metabolic Profiling: its role in biomarker discovery and gene function analysis*, Springer, 2003, pp. 111-124.
- 46
47 59. D. Turabelidze, M. Kotetishvili, A. Kreger, J. G. Morris and A. Sulakvelidze, *J. of Clin. Microbiol.*, 2000, **38**, 4242-4245.
- 48
49 60. T. L. Bannerman, G. A. Hancock, F. C. Tenover and J. M. Miller, *J. of Clin. Microbiol. y*, 1995, **33**, 551-555.
- 50
51 61. A. Alvarez-Ordóñez, D. J. M. Mouwen, M. López and M. Prieto, *J. Microbiol. Methods*, 2011, **84**, 369-378.
- 52
53 62. F. Guibet, C. Amiel, P. Cadot, C. Cordevant, M. H. Desmonts, M. Lange, A. Marecat, J. Travert, C. Denis and L. Mariey, *Vib. Spectrosc.*, 2003, **33**, 133-142.
- 54
55 63. C. Kirschner, K. Maquelin, P. Pina, N. N. Thi, L.-P. Choo-Smith, G. Sockalingum, C. Sandt, D. Ami, F. Orsini and S. Doglia, *J. of Clin. Microbiol.*, 2001, **39**, 1763-1770.
- 56
57
58
59
60

- 1
2
3 64. J. van de Vossenberg, H. Tervahauta, K. Maquelin, C. H. Blokker-Koopmans, M.
4 Uytewaal-Aarts, D. van der Kooij, A. P. van Wezel and B. van der Gaag, *Anal.*
5 *Methods*, 2013, **5**, 2679-2687.
- 6 64. U. Ch. Schroder, C. Beleites, C. Assmann, U. Glaser, U. Hubner, W. Pfister, W.
7 Fritzsche, J. Popp and U. Neugebauer, *Sci. Rep.*, 2015, **5**.
- 8 65. N. Colthup, *Introduction to infrared and Raman spectroscopy*, Elsevier, 2012.
- 9 66. N. Uzunbajakava, A. Lenferink, Y. Kraan, B. Willekens, G. Vrensen, J. Greve and
10 C. Otto, *Biopolymers*, 2003, **72**, 1-9.
- 11 67. K. Maquelin, C. Kirschner, L.-P. Choo-Smith, N. van den Braak, H. P. Endtz, D.
12 Naumann and G. Puppels, *J. Microbiol. Methods.*, 2002, **51**, 255-271.
- 13 68. W. E. Huang, M. Li, R. M. Jarvis, R. Goodacre and S. A. Banwart, in *Advances in*
14 *Applied Microbiology*, eds. I. L. Allen, S. Sima and M. G. Geoffrey, Academic
15 Press, 2010, vol. Volume 70, pp. 153-186.
- 16 69. T. M. Cotton, J. H. Kim and G. D. Chumanov, *J. Raman Spectrosc.*, 1991, **22**, 729-
17 742.
- 18 70. I. Nabiev, I. Chourpa and M. Manfait, *J. Raman Spectrosc.* 1994, **25**, 13-23.
- 19 71. R. M. Jarvis and R. Goodacre, *Chem. Soc. Rev.*, 2008, **37**, 931-936.
- 20 72. X. Shu, Y. Li, M. Liang, B. Yang, C. Liu, Y. Wang and J. Shu, *Int. J. Mass*
21 *Spectrom.*, 2012, **321-322**, 71-76.
- 22 73. R. Giebel, C. Worden, S. Rust, G. Kleinheinz, M. Robbins and T. Sandrin, *Adv. Appl.*
23 *Microbiol.*, 2010, **71**, 149-184.
- 24 74. K. Dreisewerd, *Chemical Reviews*, 2003, **103**, 395-426.
- 25 75. T. L. Williams, D. Andrzejewski, J. O. Lay and S. M. Musser, *J. Am. Soc. Mass*
26 *Spectrom.* 2003, **14**, 342-351.
- 27 76. A. Freiwald and S. Sauer, *Nat. Protoc.*, 2009, **4**, 732-742.
- 28 77. P. S. Gromski, Y. Xu, E. Correa, D. I. Ellis, M. L. Turner and R. Goodacre, *Anal.*
29 *Chim. Acta*, 2014, **829**, 1-8.
- 30 78. E. De Carolis, B. Posteraro, C. Lass-Flörl, A. Vella, A. R. Florio, R. Torelli, C.
31 Girmenia, C. Colozza, A. M. Tortorano, M. Sanguinetti and G. Fadda, *Clin.*
32 *Microbiol. Infec.*, 2012, **18**, 475-484.
- 33 79. M. Risch, D. Radjenovic, J. N. Han, M. Wydler, U. Nydegger and L. Risch, *Swiss*
34 *Med Wkly*, 2010, **140**, w13095.
- 35 80. C. Benagli, V. Rossi, M. Dolina, M. Tonolla and O. Petrini, *PLoS One*, 2011, **6**,
36 e16424.
- 37 81. T. C. Dingle and S. M. Butler-Wu, *Clin. Lab. Med.*, 2013, **33**, 589-609.
- 38 82. S. Sauer and M. Kliem, *Nature Rev.*, 2010, **8**, 74-82.
- 39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Figures

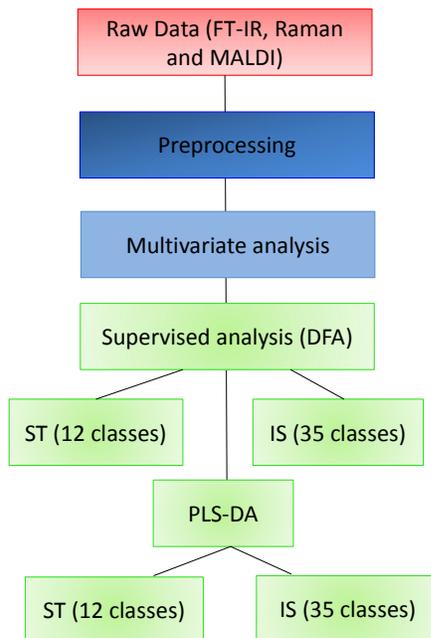


Fig 1. Workflow of data analysis undertaken for FT-IR spectroscopy, Raman spectroscopy and MALDI-TOF-MS. The data were first pre-processed then multivariate analysis MVA was applied using PC-DFA at both the (ST) strain (12 classes) and (IS) isolate (35 classes) levels. This was followed by PLS-DA.

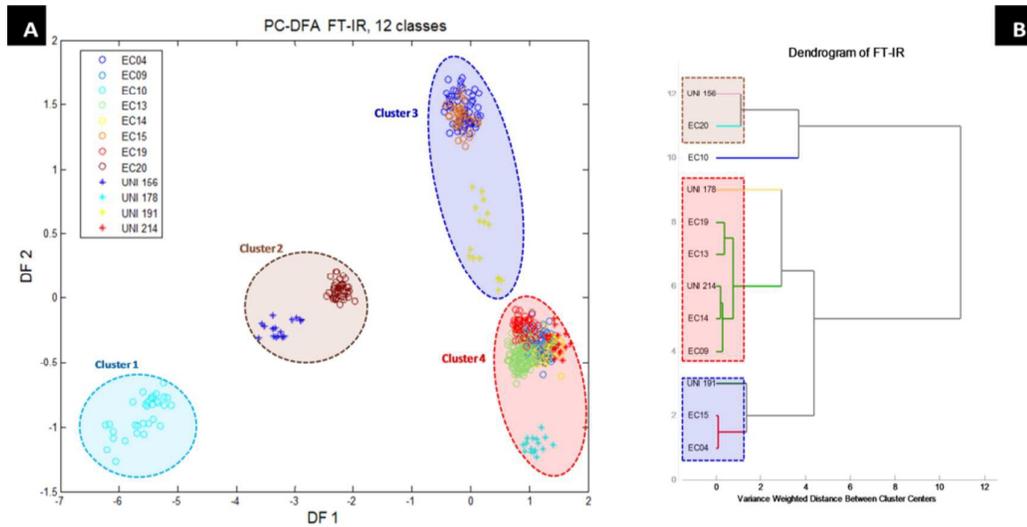


Fig 2. (A) Discriminant function analysis (DFA) scores plot from FT-IR data after pre-processing, illustrating the relationship between the 12 enterococci. (B) Cluster analysis on averaged PC-DFA scores (12 classes/strains) using Ward's linkage.

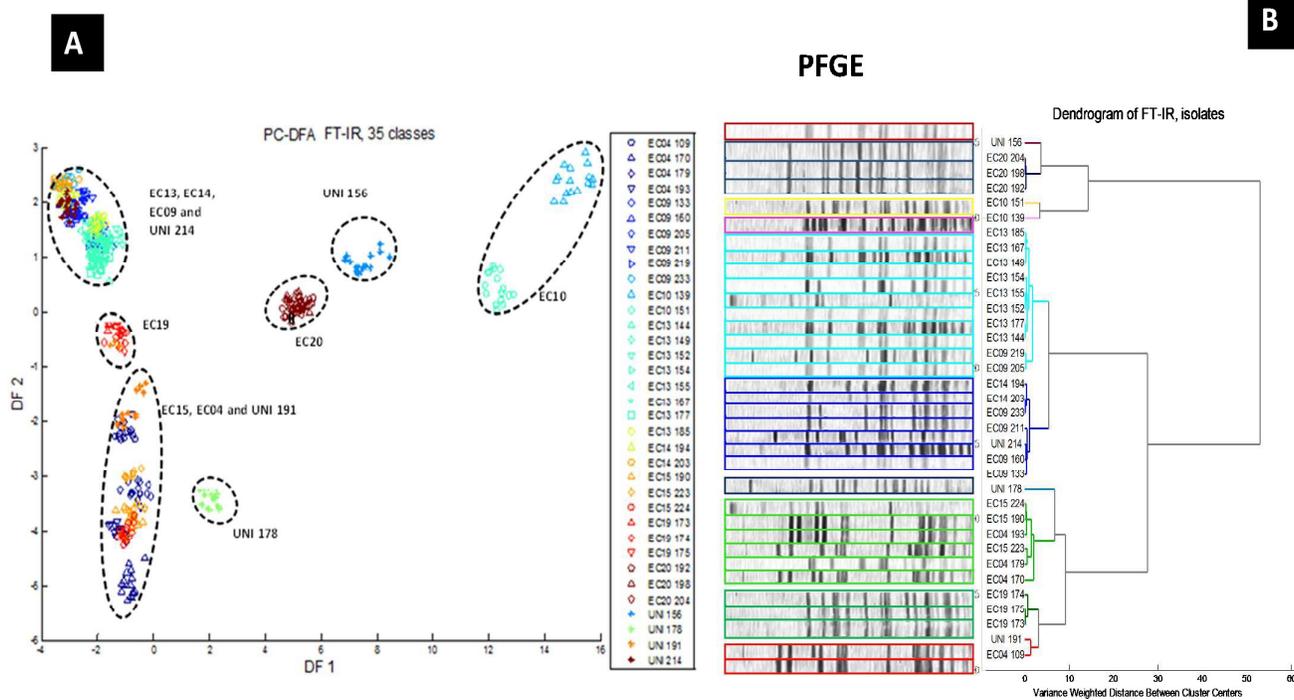


Fig 3. (A) PC-DFA plot from FT-IR data after pre-processing which illustrates the relationship between the 35 enterococcus isolates. (B) Hierarchical cluster analysis on averaged PC-DFA scores (35 classes/isolates) using Ward's linkage (right) and PFGE results (left). Each isolate is represented by the same color in both the boxes around the PFGE images and the FT-IR dendrogram.

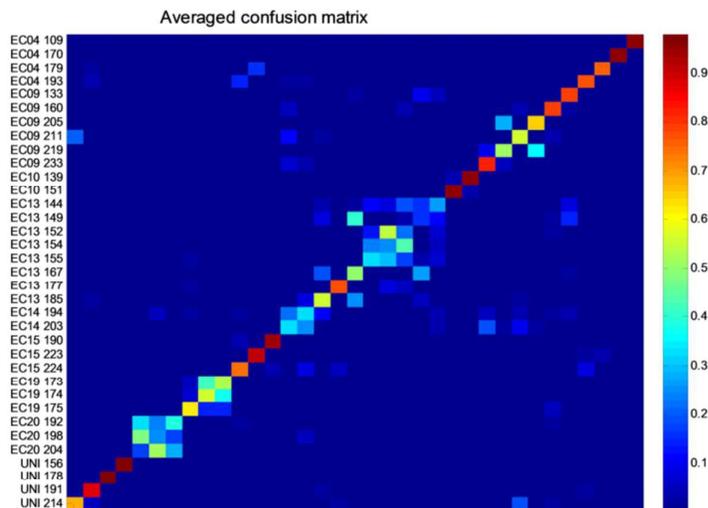


Fig 4. PLS-DA trained on 35 classes (i.e. 35 isolates) from FT-IR spectral data. High percentage class membership assignments are represented by warm colors (e.g. red) whilst the cold colors (e.g. blue) represent low percentage class membership assignments. The diagonal “tiles” are much warmer than off-diagonal “tiles”, which indicates agreement between predicted classes and known classes.

Table 1. The prediction accuracies of the 12 enterococci strains using FT-IR spectroscopy data

Class Known/Predicted	EC04	EC09	EC10	EC13	EC14	EC15	EC19	EC20	UNI 156	UNI 178	UNI 191	UNI 214
EC04	89.9%	0.5%	0.0%	0.0%	0.4%	8.3%	0.1%	0.0%	0.0%	0.0%	0.7%	0.1%
EC09	0.1%	90.3%	0.0%	1.3%	4.8%	0.0%	3.5%	0.0%	0.0%	0.0%	0.0%	0.0%
EC10	0.0%	0.1%	99.7%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%	0.1%
EC13	0.0%	0.0%	0.0%	99.8%	0.0%	0.0%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%
EC14	0.1%	48.9%	0.0%	1.1%	47.3%	1.0%	1.4%	0.1%	0.0%	0.0%	0.1%	0.0%
EC15	6.8%	1.4%	0.0%	0.0%	0.5%	91.1%	0.0%	0.0%	0.0%	0.0%	0.1%	0.0%
EC19	1.6%	9.3%	0.0%	0.2%	3.6%	0.0%	83.5%	0.0%	0.0%	0.0%	0.0%	1.8%
EC20	0.0%	0.1%	0.0%	0.0%	0.0%	0.7%	0.0%	99.2%	0.0%	0.0%	0.0%	0.0%
UNI 156	0.4%	0.0%	0.0%	0.5%	0.0%	0.0%	0.1%	0.9%	98.1%	0.0%	0.0%	0.0%
UNI 178	0.0%	5.3%	0.0%	0.1%	0.0%	0.0%	0.4%	0.0%	0.0%	93.9%	0.2%	0.0%
UNI 191	6.5%	0.9%	0.0%	25.2%	0.0%	1.3%	0.0%	0.0%	0.0%	0.0%	66.1%	0.1%
UNI 214	1.9%	13.4%	0.0%	1.0%	0.1%	0.0%	20.4%	0.0%	0.0%	0.0%	4.2%	58.9%

Table 2. The similarity between three different datasets using Procrustes distance

(A) PC-DFA at the strain level

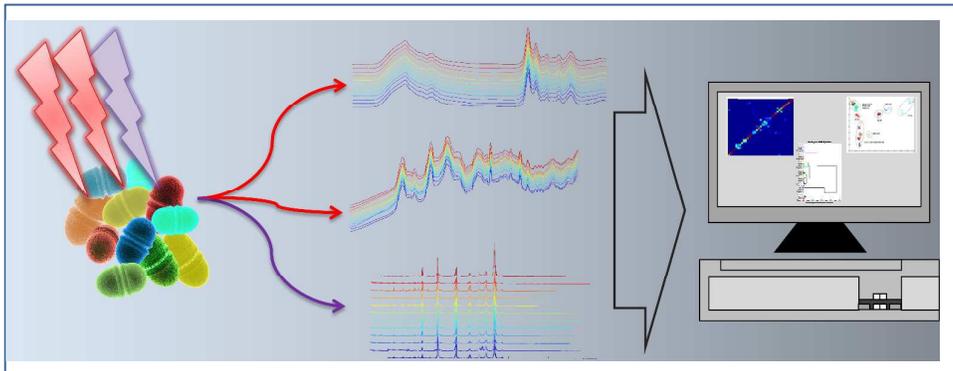
Averaging on ST level	FT-IR (IS)	FT-IR (ST)	Raman (IS)	Raman (ST)	MALDI (IS)	MALDI (ST)
FT-IR (IS)	-					
FT-IR (ST)	0.0858	-				
Raman (IS)	0.2125	0.2933	-			
Raman (ST)	0.2314	0.3187	0.1502	-		
MALDI (IS)	0.8602	0.889	0.899	0.8202	-	
MALDI (ST)	0.9125	0.8846	0.9149	0.8988	0.1812	-

(B) PC-DFA at the isolate level

Averaging on IS level	FT-IR (IS)	FT-IR (ST)	Raman (IS)	Raman (ST)	MALDI (IS)	MALDI (ST)
FT-IR (IS)	-					
FT-IR (ST)	0.1085	-				
Raman (IS)	0.2112	0.2446	-			
Raman (ST)	0.2411	0.3168	0.1132	-		
MALDI (IS)	0.8593	0.8719	0.8196	0.8001	-	
MALDI (ST)	0.8975	0.8608	0.8841	0.8703	0.0681	-

(ST) and (IS) indicate the PC-DFA was calculated at the strain (12 classes, PFGE-defined 12 types) and isolate (33 classes) levels, respectively.

TOC Graphic



1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60