Rapid discrimination of Enterococcus faecium strains using phenotypic analytical techniques

Document Version
Accepted author manuscript

[Link to publication record in Manchester Research Explorer](#)

1 **Rapid discrimination of *Enterococcus faecium* strains using phenotypic**

2 **analytical techniques and advanced chemometrics**

3

4 *Najla AlMasoud,[a] Yun Xu,[a] David I. Ellis,[a] Paul Rooney,[b] Jane F. Turton,[c] Royston*

5 *Goodacre [a #]*

6

7 *[a] School of Chemistry and Manchester Institute of Biotechnology, University of*

8 *Manchester, 131 Princess Street, Manchester, M1 7DN, UK.*

9 *[b] Belfast City Hospital, Belfast, 51 Lisburn Rd, BT9 7AB, UK.*

10 *[c] Antimicrobial Resistance and Healthcare Associated Infection Reference Unit,*

11 *National Infection Service, Public Health England, London, NW9 5EQ, UK.*

12 *[#]Correspondence to Roy Goodacre: roy.goodacre@manchester.ac.uk*

13

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.

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 meats (3), 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

3

61 techniques are sensitive and specific, they are time-consuming and their use is

62 limited by the complexity of preparation procedures and the requirement for specific

63 primers and antibodies (9-12). Nowadays, modern analytical techniques, such as

64 matrix-assisted laser desorption/ionization time-of-flight mass spectrometry

65 (MALDI-TOF-MS) (13-16), Fourier transform infrared (FT-IR) spectroscopy (17-

66 21) and Raman spectroscopy (22-24) are also used for the characterization of

67 bacteria. High dimensional and information rich datasets are produced from these

68 techniques, which has also directly led to the requirement of robust and reliable

69 chemometric methods to assist with data deconvolution and in-depth analysis (25).

70 This saw the introduction, acceptance and use of chemometrics, such as discriminant

71 function analysis (DFA) (22) and hierarchical cluster analyses (HCA) (26-28).

72 Previously, MALDI-TOF-MS has shown promising results for bacterial

73 characterization (13). FT-IR and Raman spectroscopy complement each other for

74 bacterial classification; both are robust metabolic fingerprinting techniques and need

75 little sample preparation (29, 30). FT-IR spectroscopy is used by many researchers

76 since it is not only rapid but also offers a high-throughput and non-destructive

77 method, allowing the analysis of intact bacteria and producing unique, reproducible

78 and distinct biochemical fingerprints (31). Raman spectroscopy shares similar

79 advantages to FT-IR spectroscopy and also has the additional advantage of water

80 being a very weak Raman scatter (32) so that samples do not need to be dried.

81 Here, the aim was to use these three distinct phenotypic approaches (namely

82 MALDI-TOF-MS, FT-IR and Raman spectroscopies) in combination with rigorous

83 chemometric analysis of the resultant datasets to classify 35 clinically relevant

84 isolates of enterococci, which had been previously typed by pulsed-field gel

85 electrophoresis (PFGE). This was carried out in order to compare the results from,

86 and determine the efficiency of, these analytical techniques for the rapid

87 differentiation of *E. faecium* strains. In future, this may allow clinical diagnostic

88 laboratories to analyze multiple bacterial samples rapidly for infection control

89 purposes in point-of-care setting within hospitals, clinics, or GP surgeries which

90 would significantly accelerate diagnosis, and potentially ensure that the correct

91 antimicrobial therapies were used if required, and eliminate the delay associated with

92 sending strains to reference laboratories when analyzing patient samples.

93

94

95 **MATERIALS AND METHODS**

96 **General chemicals.** Trifluoroacetic acid (TFA), HPLC grade water, acetonitrile,

97 sinapinic acid (SA), α-cyano-4-hydroxycinnamic acid (CHCA), and ferulic acid

98 (FA) were purchased from Sigma-Aldrich (Dorset, UK).

99 **Enterococci strains.** Isolates were from faecal samples from patients in a surgical

100 ward in a hospital in Belfast, UK and were collected following an increase in

101 enterococcal infections on the ward. They were identified as *E. faecium* by a

102 VITEK® system (bioMérieux) and their identity confirmed by MALDI-ToF analysis

103 using a Bruker microflex instrument. The 35 isolates were typed using pulsed-field

104 gel electrophoresis (PFGE) of *Sma*I-digested genomic DNAby Public Health

105 England's National Reference Laboratory as described previously (33). Table S1

106 summarizes information on the 35 clinical isolates, which were classified into 12

107 groups (12 PFGE-defined types) named: EC04, EC09, EC10, EC13, EC14, EC15,

108 EC19, EC20, UNI 156, UNI 178, UNI 191 and UNI 214, where 'UNI' types

109 describe isolates that were unique within the set.

110 **Bacterial isolates.**

111   The samples analyzed by the three techniques (*viz.* MALDI-TOF-MS, FT-IR and

112   Raman) were collected from the same flask to avoid any variations between different

113   preparations that may affect results obtained using the different anlaytical platforms.

114   First, enterococci were cultured on nutrient agar (NA) plates for 24 h at 37ºC. A

115   single colony from the agar culture was used to inoculate 50 mL of Lysogeny broth

116   (LB) in a 250 mL flask which was incubated overnight at 37ºC with shaking at 200

117   rpm. This was followed by measuring the optical density (OD) at 600 nm using a

118   Biomate 5 spectrophotometer (Thermo, Hemel Hempstead, UK) for each isolate.

119   The volume of analyzed bacterial suspension was then normalized to account for

120   variation in cell biomass in the different replicate cultures (4 biological replicates

121   were prepared for each isolate) and used to inoculate a fresh flask of broth, which

122   was incubated at 37ºC for 11 h (when the bacteria reached the stationary phase).

123   Then, 10 mL from each flask was collected and centrifuged at 4800 *g* for 10 min and

124   the pellet washed three times with sterile deionized water. Figure S1 illustrates the

125   preparation process.

126   For vibrational spectroscopic analysis, the collected pellets were suspended in

127   suitable volumes of saline (0.9% (w/v) NaCl) depending on the OD (all isolates had

128   approximately the same cell density). Then, 15 µL was spotted onto a silicon plate

129   (Bruker Ltd., Coventry, UK) and was allowed to dry at 40ºC for 45 min before

130   analysis with FT-IR spectroscopy. For Raman spectroscopy, 4 µL of each sample

131   was spotted onto a stainless steel plate and then allowed to dry at 40ºC for 45 min.

132   For MALDI-TOF-MS, three different matrices were tested to find the most

133   compatible matrix with enterococci; these matrices were: FA, SA and CHCA. In

134   addition, 3 different deposition methods (sample-matrix) were tested as described

135   previously (16) to find the best method for depositing the samples: mix, overlay and

6

136   underlay (data not shown). SA matrix and the mix deposition method were found to

137   be the optimal combination for MALDI-TOF-MS analysis for these samples. On the

138   day of analysis of the samples, the biomass was suspended in 1000 µL of 2% TFA

139   then vortexed for 3 min. An equal volume of 10 µL of bacterial suspension and

140   matrix were vortexed for 2 s and 2 µL of this mixture spotted onto a MALDI

141   stainless steel plate and allowed to dry at ambient temperature.

142   **Fourier transform infrared (FT-IR) spectroscopy.** FT-IR spectroscopy plate

143   (Bruker Ltd., Coventry, UK) which contained 96 locations/spots was washed using

144   5% sodium dodecyl sulfate (SDS) solution. This was followed by washing the plate

145   using deionized water and allowing it to dry at room temperature (34). High-

146   throughput screening (HTS) was carried out using a Bruker Equinox 55 FT-IR

147   spectrometer. The HTX™ module described by Winder *et al.* (35) was used with this

148   instrument. Transmission mode was used to analyze the dried biomass to produce

149   FT-IR spectra. The parameters used for FT-IR analysis included the following:

150   spectra were collected in the wavenumber range between 4000 and 600 cm^{-1},

151   resolution was 4 cm^{-1} and each spectrum was the average of 64 co-adds. Spectral

152   acquisition and subtracting the background were achieved using Opus software

153   (Bruker Ltd.). Four biological replicates, each in four analytical replicates, were

154   analyzed and analysis was performed in three machine runs, resulting in 1680 FT-IR

155   spectra.

156   **Raman Spectroscopy.** This was carried out using a confocal Raman system (inVia,

157   Renishaw plc., Wotton-Under-Edge, UK) coupled with a 785 nm wavelength laser.

158   A power intensity of ~30 mW was applied on the samples at an exposure time of

159   20 s. Four biological replicates and seven different locations within each sample spot

160   were analyzed, resulting in a total of 980 Raman spectra.

161   **MALDI-TOF-MS.** The enterococci isolates were analyzed using an AXIMA-

162   Confidence MALDI-TOF-MS (Shimadzu Biotech, Manchester, UK), equipped with

163   a nitrogen pulsed UV laser with a wavelength of 337 nm. The parameters of this

164   device were set as follows: 90 mV laser power, 91 acquired profiles with each

165   profile containing 20 shots, linear TOF, positive ionization mode, and mass-to-

166   charge (*m/z*) range of 1,000-18,000. The spectra were collected using a circular

167   raster pattern. The MALDI-TOF-MS device was calibrated using a protein mixture:

168   insulin (5,735 Da), cytochrome c (12,362 Da), and apomyoglobin (16,952 Da)

169   (Sigma-Aldrich). Each of 4 biological replicates from the 35 isolates was analyzed in

170   four technical replicates on four different days; this led to the generation of a total of

171   560 MALDI-TOF-MS spectra (35 isolates × 4 biological replicates × 4 analytical

172   replicates).

173   **Data analysis**

174   **Data pre-processing.** Opus software was used to export FT-IR data into ASCII

175   format; the data were then transferred into MATLAB 2012a (The Mathworks Inc.,

176   MA, US). All FT-IR spectra were baseline corrected using standard normal variate

177   (SNV) to remove any light scattering effect. The three analytical replicates were then

178   averaged to reduce the number of redundant samples. Due to the large number of

179   samples, 8 separate (96 spot silicon) sampling plates were used; therefore, it was

180   necessary to correct for the subtle differences in signals from different silicon plates.

181   This was achieved by using a piece-wise direct standardization (PDS) model (36).

182   The PDS model was built on two different 'reference' isolates which were spotted on

183   every plate. The pre-processed FT-IR spectra were then subjected to multivariate

184   analysis (MVA, see below). Raman spectra were also normalized using standard

185   normal variate (SNV) and then subjected to MVA.

186   MALDI-TOF-MS data were pre-processed as follows: (i) the baseline was

187   corrected using asymmetric least squares (AsLS) (37), and (ii) spectra were

188   normalized by dividing each individual baseline corrected spectrum by the square

189   root of the sum of squares of the spectrum (38). The pre-processed MALDI-TOF-

190   MS data were subjected to the same data analysis flow as Raman and FT-IR spectral

191   data.

192   **Multivariate data analysis.** A flowchart of multivariate data analysis is provided in

193   Figure 1. For all three datasets, two types of classification were performed: one at

194   the strain level (i.e. 12 classes) defined by PFGE, and the other at the isolate level

195   (i.e. 35 classes, one for each isolate).

196   For cluster analyses, principal components-discriminant function analysis (PC-

197   DFA) (39-41) was applied to reduce the dimensionality of the data and discriminate

198   samples from the designated classes. The PC-DFA scores of each class were then

199   averaged and subjected to hierarchical cluster analysis (HCA) (42). Dendrograms

200   from each analysis were generated to illustrate the relative relatedness of these

201   bacteria.

202   Partial least squares-discriminant analysis (PLS-DA) (43), with 1,000

203   bootstrapping validations (44), was also applied to obtain a validated supervised

204   classification model for discriminating different strains or isolates. In each

205   bootstrapping process, the data were randomly split into two different sets: a

206   training set and a test set. A PLS-DA model was trained on the training set and then

207   applied to the test set to predict the class membership of the samples in the test set.

208   This process was repeated 1,000 times and the results were recorded and averaged to

209   produce a $c \times c$ confusion matrix (c is the number of designated classes, either 12

210   (strains) or 35 (isolates)), in which the element at the ith row, j^{th} column is the

9

211 percentage of samples in class i being predicted as class j on average. In order to

212 assess the statistical significance of the predictive performance of the PLS-DA

213 models, a corresponding permutation test within each bootstrapping resampling was

214 also performed. This means that in addition to building the PLS-DA model using the

215 known class membership, another model (called the 'null' model) was also built

216 using a randomly permuted class membership. The results of the null model were

217 also recorded and from this the null distribution was obtained. An empirical p-value

218 was calculated by counting the number of cases where the null model had obtained

219 better predictive accuracy than the real model and dividing the obtained number by

220 the total number of bootstrapping resampling (i.e. 1,000 in this study).

221   Finally, similarities between the three different datasets (FT-IR spectroscopy,

222 Raman spectroscopy and MALDI-TOF-MS data) were measured using Procrustes

223 analysis (45). Procrustes analysis is an excellent approach for assessing the

224 differences and similarities between different ordination space from cluster analyses

225 and has been used previously for the assessment of different analytical techniques

226 (46). The distances were calculated based on the averaged PC-DFA scores for the

227 biological replicates.

228

229 **RESULTS AND DISCUSSION**

230   Table S1 shows all 35 isolates belonging to 12 strains (PFGE-defined 12 types)

231 including: EC04, EC09, EC10, EC13, EC14, EC15, EC19, EC20 UNI 156, UNI 178,

232 UNI 191 and UNI 214. These strains were previously confirmed to belong to *E.*

233 *faecium* using a VITEK® system and by MALDI-ToF analysis using a Bruker

234 Microflex system (data not shown). The PFGE results (Fig. S2) were compared to

235 results obtained in this study using FT-IR spectroscopy (17, 30, 46-49), Raman

236 spectroscopy (25, 30, 50, 51) and MALDI-TOF-MS (13, 14, 16, 52-54). We believe

237 that these analytical techniques in combination with chemometrics offer an

238 improvement in the classification of bacteria due to their higher biochemical

239 resolution.

240 **Classification using FT-IR spectroscopy.**

241 In this study, four biological replicates of bacterial isolates were analyzed in four

242 analytical replicates and analysis was performed in three machine runs, resulting in a

243 total of 1680 FT-IR spectra. The three machine replicate measurements were

244 performed in order to evaluate the reproducibility of the FT-IR technique. Typical

245 spectra based on four biological replicates of representatives of 12 strains from

246 enterococcus (EC04, EC09, EC10, EC13, EC14, EC15, EC19, EC20, UNI 156, UNI

247 178, UNI 191 and UNI 214) are provided in Figure S3A. The infrared spectra

248 contain different distinct regions that can be used to characterize bacterial samples.

249 These have been well documented previously and include: wavenumbers around

250 3400-2850 cm^{-1} corresponding to fatty acids, at 1705-1454 cm^{-1} related to amide I

251 and II regions attributed to peptides and proteins, and around 1085-1052 cm^{-1}

252 corresponding to polysaccharides (19, 55, 56).

253 Discrimination between the strains based on visual inspection of the spectra was

254 difficult (17) because these strains are very similar phenotypically. Therefore, in

255 order to develop a classification model to distinguish between bacterial samples

256 based on similarities in the spectral data, multivariate analysis was used to reduce the

257 high dimensionality of the data. First, PC-DFA was applied using 40 principal

258 components (PC) to the 12 strains (i.e. 12 classes) and 35 isolates (i.e. 35 classes)

259 using the pre-processed FT-IR spectra (Fig. 2A and 3A, respectively). Figure 2A

11

260   shows a clear separation between the 12 strains, displaying 4 main clusters; Cluster 1

261   is a single-member cluster (SMC) containing only (EC10), Cluster 2 includes (EC20

262   and UNI 156), Cluster 3 (UNI 191, EC04 and EC15) and Cluster 4 formed a large

263   group and is a combination of (EC13, EC19, EC14, EC09, UNI 214 and UNI 178).

264   Each cluster is represented by a different color in the figure. As described above,

265   HCA was undertaken using spectral data in order to simplify the DFA plot and to

266   illustrate the related strains. Cluster analysis was based on averaged DFA scores (12

267   classes/strains), using Ward's linkage as shown in Figure 2B. Clusters seen in Figure

268   2A are reflected in the HCA dendrogram plot (Fig. 2B).

269   PC-DFA was subsequently performed for all the 35 isolates and the results are

270   provided in Figure 3. Clear separation between all 35 isolates was observed despite

271   the fact that there were a much higher number of classes to be separated than the

272   number of strains. For example, clear separation was observed between the two

273   representatives of EC10 (139 and 151). Furthermore, results generated using PFGE

274   correlated well with FT-IR spectroscopic data. For example, the UNI 156 and

275   UNI 178 were seen as unique by both techniques. In addition, the three EC20

276   isolates (192, 198 and 204) and EC19 isolates (173, 174 and 175) clustered together

277   and were not differentiated using FT-IR spectroscopy, which was also observed in

278   the PFGE results, where the bands were quite similar (Fig. 3B). This implies that the

279   isolates within each of these groups are highly similar to each other phenotypically

280   and genetically. Finally, two more clusters were observed, with one cluster

281   containing all the EC04, EC15 and UNI 191 strains and the remainder of the isolates

282   forming another cluster.

283   The PLS-DA classification using FT-IR spectral data achieved an average correct

284   classification rate (CCR) of 89.4% at the strain level and 54.3% at the isolate level,

285 both with an empirical *p*-value of <0.001, i.e. not a single case where the null model

286 obtained better results, indicating that the predictive accuracies were highly

287 significant. The null distributions are provided in Figure S4A and B at the two

288 levels.

289 The confusion matrices of strains and isolates classification are presented in

290 Table 1 and Table S3, respectively. Most of the 12 strains showed high prediction

291 accuracies, for example EC04, EC10, EC13 and EC20 had accuracies of 89.9%,

292 99.7%, 99.8% and 99.2%, respectively. However, EC14 and UNI 214 had lower

293 prediction accuracies of 47.3% and 58.9%, respectively. The confusion matrix

294 showed that there was a certain level of overlap between (EC14 and EC09) and (UNI

295 214 and EC19).

296

297 Furthermore, in-depth analysis of the confusion matrix (Fig. 4) showed that

298 classification of unique strains was generally in line with PFGE results. In Figure 4,

299 high percentage class membership assignments are represented by warm colors (e.g.

300 red), indicating agreement between predicted classes and known classes. It is also

301 interesting to see that representatives from EC19 and EC20 formed two "squares" of

302 "tiles" on the diagonal line, in which the colors were similar to each other. Results

303 from Figure 4 suggest that the PLS-DA model was not able to differentiate the

304 isolates within EC19 and EC20, yet another observation that is consistent with PFGE

305 results. On the other hand, all representatives of EC04 and EC09 (160 and 133) were

306 unique in the FT-IR spectroscopy profile using the PLS-DA model but had visually

307 similar PFGE profiles. This is most likely due to PFGE providing genetic

308 information (57, 58) while FT-IR spectroscopy describes phenotypes (27, 59). This

309 implies that isolates from EC19 and EC20 are highly conserved phenotypically,

13

310 whereas those from EC04 and EC09 are not, and such subtle differences in

311 phenotypes were detected by FT-IR spectroscopy. Our observations showed that FT-

312 IR spectroscopy appears to be a very promising analytical approach for

313 discrimination of enterococci at different levels. In line with the results presented in

314 this study, work carried out by Guibet *et al.* showed that clear discrimination and

315 classification of enterococci strains can be achieved using FT-IR spectroscopy (60,

316 61).

317 **Classification using Raman spectroscopy.** In addition to the FT-IR spectroscopy

318 technique used in this study, Raman spectroscopy was used as a complementary

319 technique (17, 61-63). As expected, the two techniques generated different spectra.

320 These two approaches are complementary due to the selection rules, whereby

321 infrared causes a change in the net dipole moment in a particular functional group,

322 induced by molecular vibrations, whereas Raman causes a change in the polarization

323 of bonds within a molecule. Therefore, bonds within a molecule are generally

324 infrared or Raman active with the result being that the two techniques can provide

325 complementary (bio) chemical information (29, 64).

326 Raman spectra of the 12 *E. faecium* strains are shown in Figure S3B. Raman

327 spectra for these types appeared almost indistinguishable and no differences were

328 detected on visual inspection. Moreover, some specific peaks which were identified

329 in these spectra included: peaks at around 722 cm^{-1}, 783 cm^{-1}, 854 cm^{-1}, 1004 cm^{-1},

330 1098 cm^{-1}, 1334 cm^{-1}, 1451 cm^{-1} and 1664 cm^{-1}, which correspond to adenine,

331 cytosine/uracil, tyrosine, phenylalanine, phosphate, guanine, protein and amide I,

332 respectively (65-67).

333 PC-DFA scores plot of pre-processed Raman spectra for the 12 PFGE-defined

334 types is shown in Figure S5A. The figure shows classification results similar to those

335   seen with FT-IR spectroscopy data. There was an obvious overlap between the two

336   spectroscopic techniques, especially with representatives of EC10. However, EC20

337   overlapped with UNI 156 in FT-IR spectroscopy data, whereas EC20 was closer to

338   UNI 178 based on Raman spectroscopy data. These observations can be seen in the

339   HCA dendrogram based on Raman data (Fig. S5B), which was quite similar to the

340   HCA results generated from FT-IR data. Looking back at the dendrogram in

341   Figure S2 based on PFGE data, visual inspection showed that there were some

342   similarities between results generated via spectroscopic techniques and those based

343   on PFGE; for example, EC04 and EC15 were shown to overlap in both sets of results

344   (Fig. S2).

345   As with FT-IR data, Raman spectroscopy data on the 35 isolates were also

346   analyzed using to PC-DFA and HCA (Fig. S6A and B, respectively). The results

347   suggested that Raman spectroscopy was also successful in discriminating the two

348   representatives of EC10 (139 and 151), which was also the case using FT-IR

349   analysis (Fig. 3). Furthermore, in order to ensure the classification is robust, the data

350   were analyzed using a heat map based on PLS-DA (Fig. S6C). The results suggested

351   that all the isolates indicated as unique (UNI) by PFGE were also unique in the PLS-

352   DA model generated using Raman spectroscopy data.

353   In addition, chemometric-based identification was carried out using PLS-DA at

354   both the strain and isolate levels and the predictive accuracies were calculated based

355   on 1,000 bootstrapping resampling using Raman spectral data. The null distribution

356   was obtained (Fig. S4C and D) at both the strain (12 classes) and isolate levels (35

357   classes) resulting in average CCR of 69.3% ($p<0.001$) and 21.1% ($p<0.001$),

358   respectively. The CCR from FT-IR data was higher at both levels compared to

359   Raman data possibly due to the higher reproducibility of FT-IR data. Confusion

360   matrices were also generated at both the strain level (Table S2A) and the isolate

361   level (data not shown); these results suggested that Raman spectroscopy can also be

362   used as a robust technique for bacterial discrimination. In-depth analysis showed that

363   Raman spectroscopy generated around 70% prediction accuracy at the strain level

364   which is lower than that of FT-IR spectroscopy (nearly 90%). This is most likely due

365   to the low concentration of cells used for analysis: the infrared interrogation beam

366   used was *ca*. 1 mm and passes completely through the dried bacterial film; while the

367   Raman microscope delivers a highly focussed laser beam with an interrogation

368   volume of ~1 pL and therefore measures very few bacteria. To overcome this

369   limitation with Raman, bacteria can be analyzed directly from the agar plates or

370   surface-enhanced Raman spectroscopy (SERS) as an alternative technique (68-70),

371   but this is an area for future study.

372   **Classification using MALDI-TOF-MS.** As described in the Materials and Methods

373   section, four biological replicates were analyzed in four analytical replicates for each

374   bacterial strain, resulting in 560 MALDI-TOF-MS spectra; both the biological and

375   technical replicates clustered closely together ensuring good bioanalytical

376   reproducibility (data not shown). The spectra for the 35 enterococci isolates were

377   pre-processed before data analysis. The typical pre-processed positive ion mode

378   MALDI-TOF-MS spectra for all 12 *Enterococcus* strains (EC04, EC09, EC10,

379   EC13, EC14, EC15, EC19, EC20, UNI 156, UNI 178, UNI 191 and UNI 214) are

380   provided in Figure S3C. In general, the MALDI-TOF-MS spectra were of high

381   quality with high signal-to-noise ratios in the acquisition *m/z* range 1,000-18,000 and

382   a high number of peaks for each studied strain were detected. There are many factors

383   that can affect MALDI-TOF-MS results and some of these can differ from lab to

384   another, such as the type of medium used (71), sample handling, type of matrix (72),

385 sample deposition method (73), solvents, instrument settings (74, 75) and the type of

386 data analysis chosen (41, 76). These can inadvertently affect MALDI-TOF-MS

387 results and subsequent PC-DFA and HCA.

388 MALDI-TOF-MS spectra are not readily interpretable from the 35 isolates as

389 they are similar phenotypically and MALDI-TOF-MS spectra show only two

390 dimensions (m/z × intensity). Therefore, as is the case for the vibrational

391 spectroscopy techniques, robust multivariate analysis methods were employed for

392 this purpose. The results of PC-DFA using 12 classes (12 strains) in a three-

393 dimensional plot of DF1 *vs* DF2 *vs* DF3 and a two-dimensional plot of DF2 *vs* DF3

394 are shown in Figure S7A and B, respectively. Four main clusters were observed in

395 the PC-DFA plots; SMC (Cluster) 1 contains only UNI 178; Cluster 2 contains

396 EC20; Cluster 3 consists of EC04, EC10, EC15 and UNI 191; and Cluster 4 formed

397 a large group of (EC13, EC19, EC14, EC09, UNI 214 and UNI 156). Results from

398 the HCA dendrogram (Fig. S7C) confirmed the separation between the 12 classes

399 (i.e. 12 strains). This indicated that UNI 178 is phenotypically very different from

400 the other strains based on MALDI-TOF-MS data.

401 PC-DFA was also applied to data from the 35 isolates; the results showed

402 that isolates number 160 and 219 (both from EC09) were very different from the

403 other isolates. Therefore, another PC-DFA was carried out with these two outliers

404 removed and the HCA results are shown in Figure S8D. It appears that all

405 representatives of EC20 (204, 198 and 192) overlap with each other, which was also

406 observed in FT-IR and Raman spectroscopy data, with the exception that isolate 192

407 slightly differed from the other two representatives (204 and 198) in the HCA

408 dendrogram when using Raman data (Fig. S6B). However, analysis by PFGE

409 showed that isolates 192 and 198 clustered more closely with each other than with

410 isolate 204.

411   Furthermore, PLS-DA model applied to MALDI-TOF-MS data achieved an

412 average CCR of 78.2% ($p<0.001$) and 35.7% ($p<0.001$) for the 12 (strains) and 35

413 (isolates) classes, respectively. When PLS-DA was undertaken with 33 isolates (with

414 isolates 160 and 219 removed), the average CCR for the isolates increased to

415 53.95% ($p<0.001$). The prediction accuracies for the 12 classes are shown in Table

416 S2B and those for the 35 classes (isolates) are shown in Table S4. Table S2B shows

417 that discrimination between most of the strains (12 classes) using MALDI-TOF-MS

418 data achieved high correct classification rates, except for EC14 and UNI 191, which

419 had rather low classification rates. Confusion matrices for the 35 classes and the 33

420 classes (160 and 219 isolates removed) are shown in Figure S8A and C, respectively.

421 From these matrices, it can be seen that all the isolates identified by the reference

422 laboratory as unique (UNI), which included isolates 156, 178, 191 and 214, were

423 also classified as unique based on MALDI-TOF-MS data. Moreover, EC20 and

424 EC19 were assigned the same classification in PFGE typing, and this was in

425 agreement with MALDI-TOF-MS, FT-IR spectroscopy and Raman spectroscopy

426 data. In addition, based on MALDI-TOF-MS data (Fig.S8A and C), representatives

427 of EC13 (152, 154 and 155) belonged to the same cluster, and isolates 177 from

428 EC13 was significantly different from the remaining EC13 strains; this was also

429 observed in FT-IR and PFGE data. Looking back at Figure S8C, it can be seen that

430 all the strains from EC04 were unique in MALDI-TOF-MS and FT-IR profiles when

431 using PLS-DA modelling.

432 **Procrustes distance test of the three analytical techniques.** Analytical techniques

433 such as FT-IR spectroscopy, Raman spectroscopy and MALDI-TOF-MS are

434   currently used in clinical research studies worldwide and many reports have been

435   published showing advantages of using such techniques (24, 54, 77, 78). Kirschner

436   *et al*. (61) demonstrated accurate identification and classification of 18 strains from 6

437   different species belonging to enterococci using vibrational spectroscopic techniques

438   in combination with chemometrics. This study suggested that FT-IR and Raman

439   spectroscopies can offer potential alternatives to the conventional typing tests due to

440   their speed and ease of use. Oliveira *et al*. (51) showed that Raman spectroscopy, in

441   combination with a chemometric algorithm, can be used to discriminate between

442   seven different colonies of Gram-positive and Gram-negative bacteria. In another

443   previous study, it was also shown that 59 clinical bacterial strains associated with

444   urinary tract infections (UTIs) could be identified using FT-IR and Raman

445   spectroscopy (17). As an alternative to vibrational spectroscopic techniques,

446   MALDI-TOF-MS is a relatively new technique which has shown very promising

447   results for identification in agreement with methodologies carried out in

448   microbiological laboratories, and therefore has been used for the identification and

449   classification of bacterial species (15, 79, 80) and is appearing in many clinical

450   microbiology testing laboratories (54, 81, 82).

451   Previous studies have generally focussed on the application of just one or two

452   analytical techniques for the classification of *Enterococcus* spp. However, to

453   generate complementary data and more comprehensive analysis, this study combines

454   three different analytical techniques – FT-IR spectroscopy, Raman spectroscopy and

455   MALDI-TOF-MS – to analyze whole bacterial cells. Successful classification was

456   demonstrated at the strain (i.e. 12 classes) and isolate (i.e. 35 classes) levels based on

457   data generated by the three analytical platforms. In order to assess the overall

458   information content in the spectra that has been revealed by the cluster analysis from

459   the scores plots, Procrustes analysis was employed to assess the overall similarity

460   between the patterns detected by these three platforms. The results are presented in

461   terms of Procrustes distance (Table 2A and B), where the Procrustes distance varies

462   from 0 to 1; the lower the distance, the higher the similarity between the results. The

463   comparisons were made using averaged PC-DFA scores. For each dataset, there

464   were two sets of PC-DFA scores, one at the strain level (12 classes) and another for

465   isolates classification (33 classes). For each set of PC-DFA scores, the scores were

466   then averaged according to their strain label and isolate label to give two sets of

467   *averaged* PC-DFA scores.

468   The findings in Table 2 can be summarized as follows:

469   (i)   The patterns in the PC-DFA scores at strain and isolate levels were highly

470   similar to each other for all the three analytical platforms. The Procrustes

471   distances varied from 0.0681 to 0.1812. This suggested that the variation

472   originating from different bacteria is the main factor in PC-DFA, i.e. the

473   differences between different bacterial genotypes were significantly higher

474   than those between different isolates.

475   (ii)   The two vibrational spectroscopic techniques (FT-IR and Raman) generated

476   highly similar results both at the strain and isolate classification levels, with

477   the corresponding Procrustes distances varying from 0.2112 to 0.3187.

478   (iii)   However, the results generated by MALDI-TOF-MS were significantly

479   different from those generated by the two spectroscopic techniques, and the

480   corresponding Procrustes distances were all above 0.8. Such differences can

481   be mainly attributed to data on isolate UNI 178, which appeared to be very

482   different to other strains in the MALDI-TOF-MS dataset.

483   Table S5 shows a summative comparison of the 4 main clusters identified based on

484   the three analytical techniques using PC-DFA plots of the 12 *E. faecium* strains (12

485   classes). It can be seen from this table that despite the large Procrustes distances

486   between data generated by MALDI-TOF-MS and those generated by the other two

487   techniques, the main identified clusters patterns observed in all three datasets were

488   still largely consistent.

489

490

491   **CONCLUSIONS**

492   The results obtained from the two vibrational spectroscopic techniques

493   demonstrated that good discrimination can be achieved at both the strain and isolate

494   levels and the detected patterns from the two techniques were highly similar. In

495   addition, bacterial classification results from MALDI-TOF-MS were generally

496   consistent with these vibrational spectroscopic techniques. However, UNI 178 was

497   detected to be very different in MALDI-TOF-MS data, which differed from the other

498   two analytical techniques employed in this study.

499   The results obtained using these spectroscopic phenotyping approaches were

500   mostly consistent with previous results obtained from experiments carried out using

501   the genotypic classification method, PFGE. Some of the results differed when

502   directly comparing our analytical approach with results from the molecular approach

503   and these differences may be due to comparing phenotypic differences from whole-

504   organism fingerprinting with genotypic differences using PFGE.

505   In conclusion, we have assessed multiple analytical phenotypic as

506   complementary approaches to current molecular methods. All methods provided

507   excellent clustering which was in general agreement with genotypic baseline

508 methods, as well as allowing excellent discrimination to the strain level and good

509 resolution at the sub-strain level. We believe that these three different

510 physicochemical techniques have excellent potential as high-throughput point-of-

511 care screening tools, and for the rapid and reproducible classification of clinically

512 relevant bacteria, such as *E. faecium*.

513 **ACKNOWLEDGMENTS**

518

REFERENCES

1. **Schleifer KH, Kilpper-Bälz R.** 1984. Transfer of *Streptococcus faecalis* and *Streptococcus faecium* to the Genus *Enterococcus* nom. rev. as *Enterococcus faecalis* comb. nov. and *Enterococcus faecium* comb. nov. Int. J. Syst Bacteriol. **34:**31-34.

2. **Kayser FH, Bienz KA, Eckert J.** 2011. Medical microbiology. Thieme.

3. **Hayes JR, English LL, Carter PJ, Proescholdt T, Lee KY, Wagner DD, White DG.** 2003. Prevalence and antimicrobial resistance of *Enterococcus* species straind from retail meats. Appl. Environ. Microbiol. **69:**7153-7160.

4. **Franz CM, Stiles ME, Schleifer KH, Holzapfel WH.** 2003. Enterococci in foods a conundrum for food safety. Int. J. Food Microbiol. **88:**105-122.

5. **McCracken M, Wong A, Mitchell R, Gravel D, Conly J, Embil J, Johnston L, Matlow A, Ormiston D, Simor A.** 2013. Molecular epidemiology of vancomycin-resistant enterococcal bacteraemia: results from the Canadian Nosocomial Infection Surveillance Program, 1999–2009. J. Antimicrob. Chemother. **68**:1505-9**.**

6. **Woodford N.** 1998. Glycopeptide-resistant enterococci: a decade of experience. J. Med. Microbiol. **47:**849-862.

7. **Arias CA, Murray BE.** 2012. The rise of the Enterococcus: beyond vancomycin resistance. Nature Rev. Microbiol. **10:**266-278.

8. **Altekruse S, Cohen M, Swerdlow D.** 1997. Emerging foodborne diseases. Emerg. Infect. Dis. **3:**285-294.

9. **Engvall E.** 1977. Quantitative enzyme immunoassay (ELISA) in microbiology. Med. Biol. **55:**193.

10. **Yolken RH.** 1980. Enzyme-linked immunosorbent assay (ELISA): a practical tool for rapid diagnosis of viruses and other infectious agents. Yale J. Biol. Med. and medicine **53:**85.

545   11.   **Ke D, Picard FJ, Martineau F, Ménard C, Roy PH, Ouellette M, Bergeron MG.**

546   1999. Development of a PCR assay for rapid detection of enterococci. J. Clin.

547   Microbiol. **37:**3497-3503.

548   12.   **Reen DJ.** 1994. Enzyme-linked immunosorbent assay (ELISA), p 461-466, Basic

549   Protein and Peptide Protocols. Springer.

550   13.   **Claydon MA, Davey SN, Edwards-Jones V, Gordon DB.** 1996. The rapid

551   identification of intact microorganisms using mass spectrometry. Nat. Biotechnol.

552   **14:**1584-1586.

553   14.   **Quintela-Baluja M, Böhme K, Fernández-No IC, Morandi S, Alnakip ME,**

554   **Caamaño-Antelo S, Barros-Velázquez J, Calo-Mata P.** 2013. Characterization of

555   different food-straind *Enterococcus* strains by MALDI-TOF mass fingerprinting.

556   Electrophoresis **34:**2240-2250.

557   15.   **Lasch P, Fleige C, Stämmler M, Layer F, Nübel U, Witte W, Werner G.** 2014.

558   Insufficient discriminatory power of MALDI-TOF mass spectrometry for typing of

559   *Enterococcus faecium* and *Staphylococcus aureus* strains. J. Microbiol. Methods

560   **100:**58-69.

561   16.   **AlMasoud N, Xu Y, Nicolaou N, Goodacre R.** 2014. Optimization of matrix

562   assisted desorption/ionization time of flight mass spectrometry (MALDI-TOF-MS)

563   for the characterization of *Bacillus* and *Brevibacillus* species. Anal. Chim. Acta.

564   **840:**49-57.

565   17.   **Goodacre R, Burton R, Kaderbhai N, Woodward AM, Kell DB, Rooney PJ.**

566   1998. Rapid identification of urinary tract infection bacteria using hyperspectral

567   whole-organism fingerprinting and artificial neural networks. Microbiology

568   **144:**1157-1170.

569   18.   **Helm D, Labischinski H, Schallehn G, Naumann D.** 1991. Classification and

570   identification of bacteria by Fourier-transform infrared spectroscopy. J. Gen.

571   Microbiol. **137:**69-79.

572   19.   **Naumann D, Helm D, Labischinski H.** 1991. Microbiological characterizations by

573   FT-IR spectroscopy. Nature **351:**81-82.

574   20.   **Naumann D.** 2000. Infrared spectroscopy in microbiology. Encyclopedia of

575   analytical chemistry. John Wiley and Sons Ltd, Chichester.

576   21.   **Burgula Y, Khali D, Kim S, Krishnan SS, Cousin MA, Gore JP, Reuhs BL,**

577   **Mauer LJ.** 2007. Review of mid-infrared fourier transform-infrared spectroscopy

578   applications for bacterial detection. J. Rapid Methods Autom. Microbiol. **15:**146-

579   175.

580   22.   **López-Díez EC, Goodacre R.** 2004. Characterization of microorganisms using UV

581   resonance Raman spectroscopy and chemometrics. Anal. chem. **76:**585-591.

582   23.   **Maquelin K, Kirschner C, Choo-Smith LP, van den Braak N, Endtz HP,**

583   **Naumann D, Puppels GJ.** 2002. Identification of medically relevant

584   microorganisms by vibrational spectroscopy. J. Microbiol. Methods **51:**255-271.

585   24.   **Beekes M, Lasch P, Naumann D.** 2007. Analytical applications of Fourier

586   transform-infrared (FT-IR) spectroscopy in microbiology and prion research. Vet.

587   Microbiol. **123:**305-319.

588   25.   **Ellis DI, Cowcher DP, Ashton L, O'Hagan S, Goodacre R.** 2013. Illuminating

589   disease and enlightening biomedicine: Raman spectroscopy as a diagnostic tool.

590   Analyst **138:**3871-3884.

591   26.   **Gutteridge CS, Valus L, Macfie HJH.** 1985. 14 - Numerical Methods in the

592   Classification of Micro-organisms by Pyrolysis Mass Spectrometry, p 369-401. *In*

593   Priest MGJG (ed), Computer-Assisted Bacterial Systematics. Academic Press,

594   London.

595   27.   **Davis R, Mauer L.** 2010. Fourier transform infrared (FT-IR) spectroscopy: a rapid

596   tool for detection and analysis of foodborne pathogenic bacteria. Current research,

597   technology and education topics in applied microbiology and microbial

598   biotechnology **2:**1582-1594.

599 28. **Dworzanski JP, Deshpande SV, Chen R, Jabbour RE, Snyder AP, Wick CH, Li**

600   **L.** 2006. Mass Spectrometry-Based Proteomics Combined with Bioinformatic Tools

601   for Bacterial Classification. J. Proteome. Res. **5:**76-87.

602 29. **Goodacre R, Radovic BS, Anklam E.** 2002. Progress toward the rapid

603   nondestructive assessment of the floral origin of European honey using dispersive

604   Raman spectroscopy. Appl. Spectrosc. **56:**521-527.

605 30. **Ellis DI, Goodacre R.** 2006. Metabolic fingerprinting in disease diagnosis:

606   biomedical applications of infrared and Raman spectroscopy. Analyst **131:**875-885.

607 31. **Argyri AA, Jarvis RM, Wedge D, Xu Y, Panagou EZ, Goodacre R, Nychas G-**

608   **JE.** 2013. A comparison of Raman and FT-IR spectroscopy for the prediction of

609   meat spoilage. Food Control **29:**461-470.

610 32. **Goodacre R, Timmins EM, Rooney PJ, Rowland JJ, Kell DB.** 1996. Rapid

611   identification of *Streptococcus* and *Enterococcus* species using diffuse reflectance-

612   absorbance Fourier transform infrared spectroscopy and artificial neural networks.

613   FEMS Microbiol. Lett. **140:**233-239.

614 33. **Woodford N, Reynolds R, Turton J, Scott F, Sinclair A, Williams A, Livermore D**.

615   2003. Two widely disseminated strains of Enterococcus faecalis highly resistant to

616   gentamicin and ciprofloxacin from bacteraemias in the UK and Ireland. J

617   Antimicrob Chemother. **52**:711-714.

618 34. **Patel SA, Currie F, Thakker N, Goodacre R.** 2008. Spatial metabolic

619   fingerprinting using FT-IR spectroscopy: investigating abiotic stresses on

620   Micrasterias hardyi. Analyst **133:**1707-1713.

621 35. **Winder CL, Gordon SV, Dale J, Hewinson RG, Goodacre R.** 2006. Metabolic

622   fingerprints of Mycobacterium bovis cluster with molecular type: implications for

623   genotype–phenotype links. Microbiology **152:**2757-2765.

624 36. **Wang Y, Veltkamp DJ, Kowalski BR.** 1991. Multivariate instrument

625   standardization. Anal. Chem. **63:**2750-2756.

626 37. **Eilers PHC.** 2004. Parametric Time Warping. Anal. Chem. **76:**404-411.

627 38. **Brereton RG.** 2003. Chemometrics: data analysis for the laboratory and chemical

628 plant. John Wiley and Sons.

629 39. **Manly BF.** 2004. Multivariate statistical methods. a primer. CRC Press.

630 40. **Harrigan GG, LaPlante RH, Cosma GN, Cockerell G, Goodacre R, Maddox**

631 **JF, Luyendyk JP, Ganey PE, Roth RA.** 2004. Application of high-throughput

632 Fourier-transform infrared spectroscopy in toxicology studies: contribution to a

633 study on the development of an animal model for idiosyncratic toxicity. Toxicol.

634 Lett. **146:**197-205.

635 41. **Gromski PS, Muhamadali H, Ellis DI, Xu Y, Correa E, Turner ML, Goodacre**

636 **R.** 2015. A tutorial review: Metabolomics and partial least squares-discriminant

637 analysis–a marriage of convenience or a shotgun wedding. Anal. Chim. Act.

638 **879:**10-23.

639 42. **Hastie T, Tibshirani R, Friedman J.** 2009. The Elements of Statistical Learnin.

640 Springer, New York.

641 43. **Barker M, Rayens W.** 2003. Partial least squares for discrimination. J. Chemom.

642 **17:**166-173.

643 44. **Efron B, Tibshirani RJ.** 1994. An introduction to the bootstrap. Chapman and

644 Hall/CRC press.

645 45. **Gower JC, Dijksterhuis GB.** 2004. Procrustes problems, vol 3. Oxford University

646 Press Oxford.

647 46. **AlRabiah H, Xu Y, Rattray NJW, Vaughan AA, Gibreel T, Sayqal A, Upton M,**

648 **Allwood JW, Goodacre R.** 2014. Multiple metabolomics of uropathogenic E. coli

649 reveal different information content in terms of metabolic potential compared to

650 virulence factors. Analyst **139:**4193-4199.

651 47. **Naumann D.** 1984. Some ultrastructural information on intact, living bacterial cells

652 and related cell-wall fragments as given by FTIR. Infrared Phys. **24:**233-238.

653 48. **Naumann D, Fijala V, Labischinski H, Giesbrecht P.** 1988. The rapid

654 differentiation and identification of pathogenic bacteria using Fourier transform

655   infrared spectroscopic and multivariate statistical analysis. J. Mol. Struct. **174:**165-
656   170.

657   49.   **Mariey L, Signolle JP, Amiel C, Travert J.** 2001. Discrimination, classification,
658   identification of microorganisms using FTIR spectroscopy and chemometrics. Vib.
659   Spectrosc. **26:**151-159.

660   50.   **Muhamadali H, Chisanga M, Subaihi A, Goodacre R.** 2015. Combining Raman
661   and FT-IR Spectroscopy with Quantitative Isotopic Labeling for Differentiation of
662   *E. coli* Cells at Community and Single Cell Levels. Anal. Chem. **87:**4578-4586.

663   51.   **de Siqueira e Oliveira FS, Giana HE, Silveira JL.** 2012. Discrimination of
664   selected species of pathogenic bacteria using near-infrared Raman spectroscopy and
665   principal components analysis. J. Biomed. Opt. **17:**107004

666   52.   **Cramer R, Gobom J, Nordhoff E.** 2005. High-throughput proteomics using
667   matrix-assisted laser desorption/ionization mass spectrometry. Expert Rev.
668   Proteomics. **2**: 407-20

669   53.   **Ellis DI, Dunn WB, Griffin JL, Allwood JW, Goodacre R.** 2007. Metabolic
670   fingerprinting as a diagnostic tool. Analyst, **131:**875-885.

671   54.   **Carbonnelle E, Mesquita C, Bille E, Day N, Dauphin B, Beretti J-L, Ferroni A,**
672   **Gutmann L, Nassif X.** 2011. MALDI-TOF mass spectrometry tools for bacterial
673   identification in clinical microbiology laboratory. Clin. Biochem. **44:**104-109.

674   55.   **Kim S, Reuhs BL, Mauer LJ.** 2005. Use of Fourier transform infrared spectra of
675   crude bacterial lipopolysaccharides and chemometrics for differentiation of
676   Salmonella enterica serotypes. J. Appl. Microbiol. **99:**411-417.

677   56.   **Ellis DI, Harrigan GG, Goodacre R.** 2003. Metabolic fingerprinting with Fourier
678   transform infrared spectroscopy, Metabolic Profiling: its role in biomarker discovery
679   and gene function analysis. p 111-124, Springer. Norwell, Massachusetts, USA.

680   57.   **Turabelidze D, Kotetishvili M, Kreger A, Morris JG, Sulakvelidze A.** 2000.
681   Improved pulsed-field gel electrophoresis for typing vancomycin-resistant
682   enterococci. J. Clin. Microbiol. **38:**4242-4245.

683   58.   **Bannerman TL, Hancock GA, Tenover FC, Miller JM.** 1995. Pulsed-field gel
684   electrophoresis as a replacement for bacteriophage typing of *Staphylococcus aureus*.
685   J. Clin. Microbiol. **33:**551-555.

686   59.   **Alvarez-Ordóñez A, Mouwen DJM, López M, Prieto M.** 2011. Fourier transform
687   infrared spectroscopy as a tool to characterize molecular composition and stress
688   response in foodborne pathogenic bacteria. J. Microbiol. Meth. **84:**369-378.

689   60.   **Guibet F, Amiel C, Cadot P, Cordevant C, Desmonts MH, Lange M, Marecat
690   A, Travert J, Denis C, Mariey L.** 2003. Discrimination and classification of
691   Enterococci by Fourier transform infrared (FT-IR) spectroscopy. Vib. Spectrosc.
692   **33:**133-142.

693   61.   **Kirschner C, Maquelin K, Pina P, Thi NN, Choo-Smith L-P, Sockalingum G,
694   Sandt C, Ami D, Orsini F, Doglia S.** 2001. Classification and identification of
695   enterococci: a comparative phenotypic, genotypic, and vibrational spectroscopic
696   study. J. Clin. Microbiol. **39:**1763-1770.

697   62.   **Van de Vossenberg J, Tervahauta H, Maquelin K, Blokker-Koopmans CH,
698   Uytewaal-Aarts M, van der Kooij D, van Wezel AP, van der Gaag B.** 2013.
699   Identification of bacteria in drinking water with Raman spectroscopy. Anal. Meth.
700   **5:**2679-2687.

701   63.   **Ch. Schroder U, Beleites C, Assmann C, Glaser U, Hubner U, Pfister W,
702   Fritzsche W, Popp J, Neugebauer U.** 2015. Detection of vancomycin resistances
703   in enterococci within 3 [1/2] hours. Sci. Rep. **5:**8217.

704   64.   **Colthup N.** 2012. Introduction to infrared and Raman spectroscopy. New York,
705   Elsevier.

706   65.   **Uzunbajakava N, Lenferink A, Kraan Y, Willekens B, Vrensen G, Greve J,
707   Otto C.** 2003. Nonresonant Raman imaging of protein distribution in single human
708   cells. Biopolymers **72:**1-9.

709   66.   **Maquelin K, Kirschner C, Choo-Smith L-P, van den Braak N, Endtz HP, Naumann D, Puppels G.** 2002. Identification of medically relevant microorganisms by vibrational spectroscopy. J. Microbiol. Methods **51:**255-271.

712   67.   **Huang WE, Li M, Jarvis RM, Goodacre R, Banwart SA.** 2010. Shining Light on the Microbial World: The Application of Raman Microspectroscopy. Adv. Appl. Microbiol. **70:** 931-936.

715   68.   **Cotton TM, Kim JH, Chumanov GD.** 1991. Application of surface-enhanced Raman spectroscopy to biological systems. J. Raman Spectrosc. **22:**729-742.

717   69.   **Nabiev I, Chourpa I, Manfait M.** 1994. Applications of Raman and surface-enhanced Raman scattering spectroscopy in medicine. J. Raman Spectrosc. **25:**13-23.

720   70.   **Jarvis RM, Goodacre R.** 2008. Characterisation and identification of bacteria using SERS. Chem. Soc. Rev. **37:**931-936.

722   71.   **Shu X, Li Y, Liang M, Yang B, Liu C, Wang Y, Shu J.** 2012. Rapid lipid profiling of bacteria by online MALDI-TOF mass spectrometry. nt. J. Mass spectrom. **321–322:**71-76.

725   72.   **Giebel R, Worden C, Rust SM, Kleinheinz GT, Robbins M, Sandrin TR**. 2010. Chapter 6 - Microbial Fingerprinting using Matrix-Assisted Laser Desorption Ionization Time-Of-Flight Mass Spectrometry (MALDI-TOF MS): Applications and Challenges, p149-184, Advances in Applied Microbiology, vol Volume 71. Academic Press.

730   73.   **Dreisewerd K.** 2003. The desorption process in MALDI. Chemical reviews **103:**395-426.

732   74.   **Williams TL, Andrzejewski D, Lay JO, Musser SM.** 2003. Experimental factors affecting the quality and reproducibility of MALDI TOF mass spectra obtained from whole bacteria cells. J. Am. Soc. Mass Spectrom. **14:**342-351.

735   75.   **Freiwald A, Sauer S.** 2009. Phylogenetic classification and identification of bacteria by mass spectrometry. Nat. Protoc **4:**732-742.

737 76. **Gromski PS, Xu Y, Correa E, Ellis DI, Turner ML, Goodacre R.** 2014. A
738   comparative investigation of modern feature selection and classification approaches
739   for the analysis of mass spectrometry data. Anal. Chim. Act. **829:**1-8.

740 77. **De Carolis E, Posteraro B, Lass-Flörl C, Vella A, Florio AR, Torelli R,**
741   **Girmenia C, Colozza C, Tortorano AM, Sanguinetti M, Fadda G.** 2012. Species
742   identification of Aspergillus, Fusarium and Mucorales with direct surface analysis
743   by matrix-assisted laser desorption ionization time-of-flight mass spectrometry.
744   Clin. Microbiol. Infec. **18:**475-484.

745 78. **Risch M, Radjenovic D, Han JN, Wydler M, Nydegger U, Risch L.** 2010.
746   Comparison of MALDI TOF with conventional identification of clinically relevant
747   bacteria. Swiss Med. Wkly. **140:**w13095.

748 79. **Benagli C, Rossi V, Dolina M, Tonolla M, Petrini O.** 2011. Matrix-assisted laser
749   desorption ionization-time of flight mass spectrometry for the identification of
750   clinically relevant bacteria. PLoS One **6:**e16424.

751 80. **Dingle TC, Butler-Wu SM.** 2013. MALDI-TOF mass spectrometry for
752   microorganism identification. Clin. Lab. Med. **33:**589-609.

753 81. **Sauer S, Kliem M.** 2010. Mass spectrometry tools for the classification and
754   identification of bacteria. Nat Rev Micro **8:**74-82.

755 82. **Bizzini A, Durussel C, Bille J, Greub G, Prod'hom G.** 2010. Performance of
756   matrix-assisted laser desorption ionization-time of flight mass spectrometry for
757   identification of bacterial strains routinely isolated in a clinical microbiology
758   laboratory. J Clin Microbiol **48:**1549-1554.

FIGURE LEGENDS

760 **FIG 1.** Workflow of data analysis undertaken for FT-IR spectroscopy, Raman spectroscopy

761 and MALDI-TOF-MS. The data were first pre-processed then multivariate analysis MVA

762 was applied using PC-DFA at both the (ST) strain (12 classes) and (IS) isolate (35 classes)

763 levels. This was followed by PLS-DA.

764 **FIG 2.** (A) Discriminant function analysis (DFA) scores plot from FT-IR data after pre-

765 processing, illustrating the relationship between the 12 enterococci. (B) Cluster analysis on

766 averaged PC-DFA scores (12 classes/strains) using Ward's linkage.

767 **FIG 3.** (A) PC-DFA plot from FT-IR data after pre-processing which illustrates the

768 relationship between the 35 enterococcus isolates. (B) Hierarchical cluster analysis on

769 averaged PC-DFA scores (35 classes/isolates) using Ward's linkage (right) and PFGE

770 results (left). Each isolate is represented by the same color in both the boxes around the

771 PFGE images and the FT-IR dendrogram.

772 **FIG 4.** PLS-DA trained on 35 classes (i.e. 35 isolates) from FT-IR spectral data. High

773 percentage class membership assignments are represented by warm colors (e.g. red) whilst

774 the cold colors (e.g. blue) represent low percentage class membership assignments. The

775 diagonal "tiles" are much warmer than off-diagonal "tiles", which indicates agreement

776 between predicted classes and known classes.
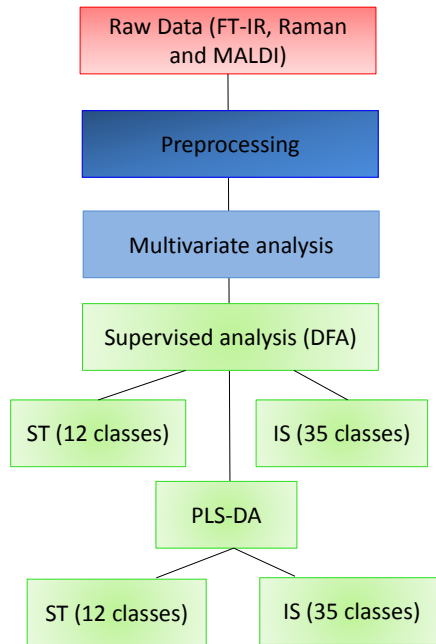
777

778   Figure 1
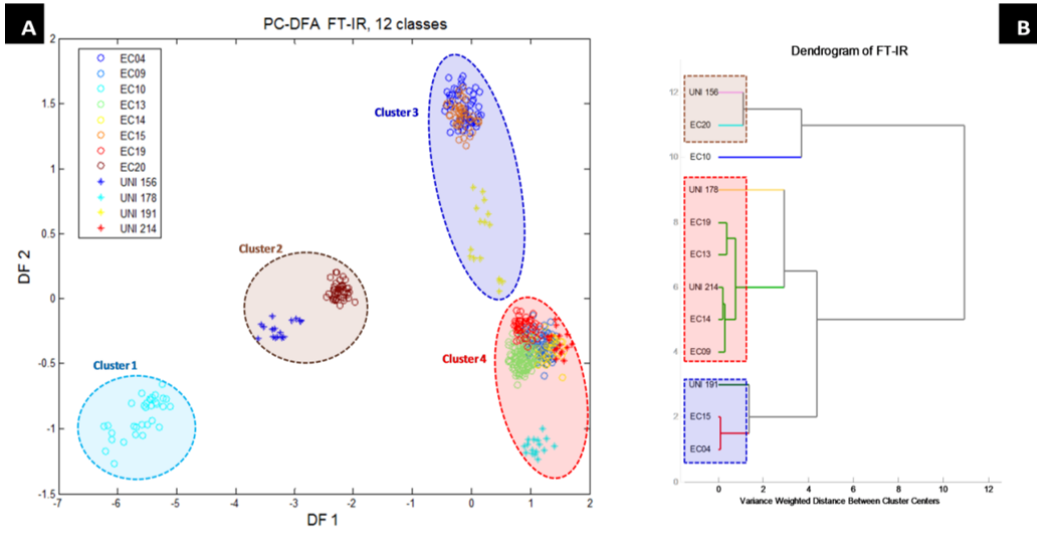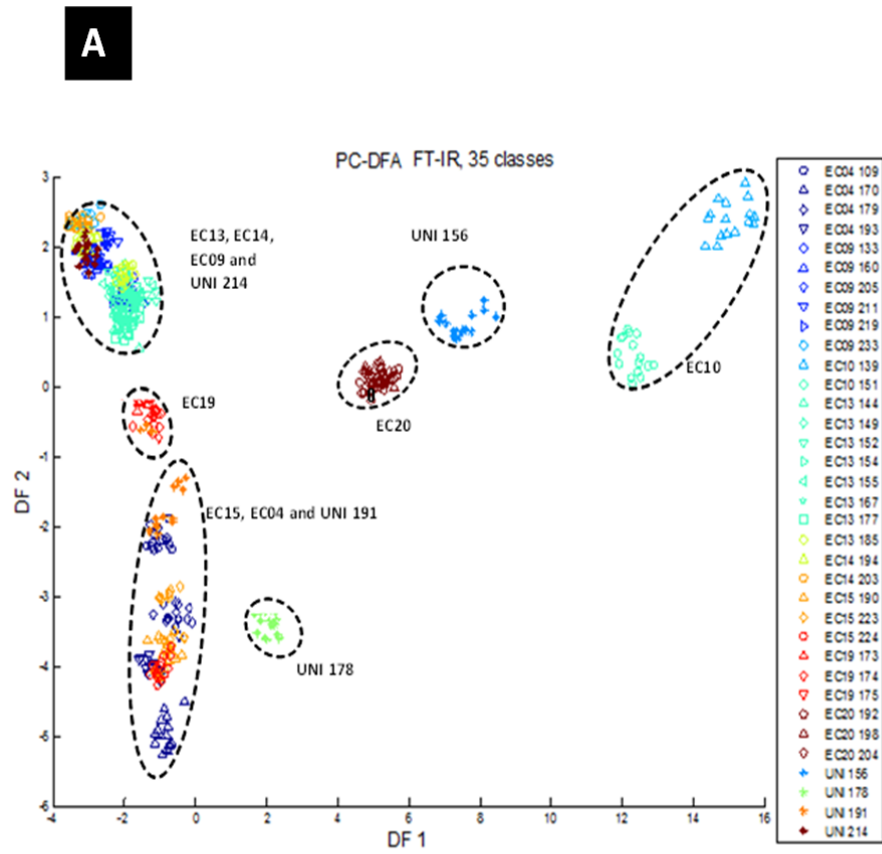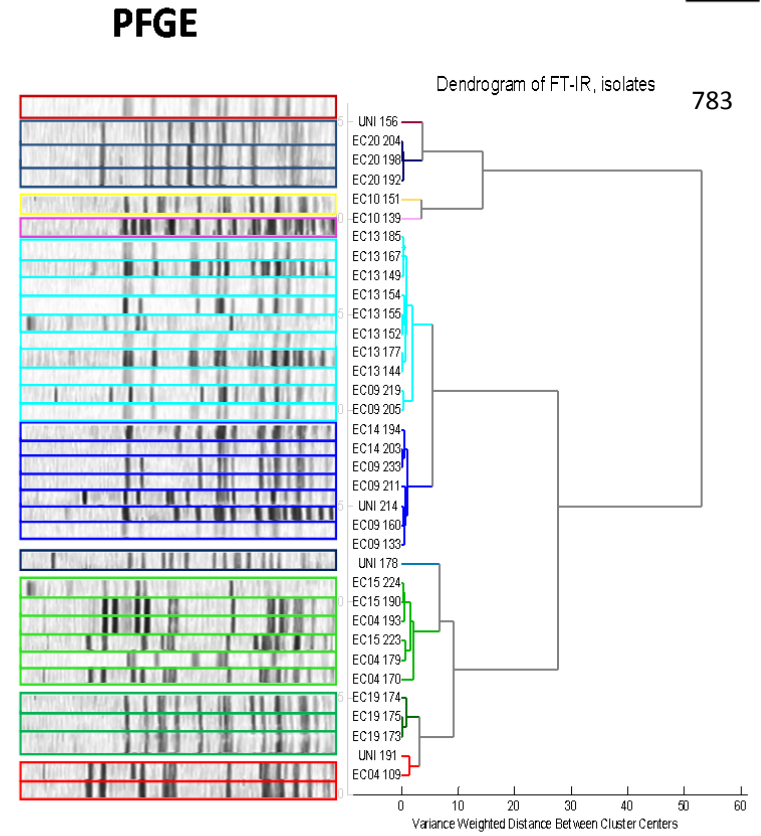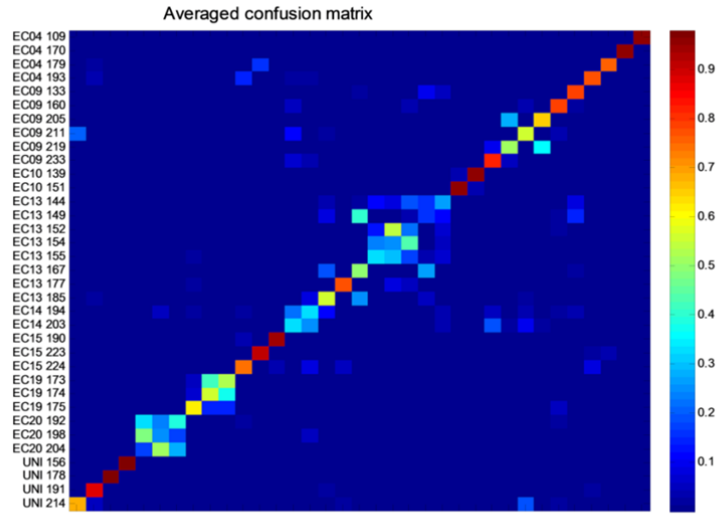
779

780   Figure 2

781

782

Figure 3

35

Averaged confusion matrix

785

786   Figure 4

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

788 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%

789

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

791 (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	-

792

793 (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	-

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